Requirements Engineering and Tool-Support for Security and Privacy

Dr. rer. nat. Sebastian Pape

Habilitation thesis submitted in
fulfillment of the requirements for
the academic title Dr. habil.
(Doctor habilitatus)

Submitted to the Faculty of Computer Science and Mathematics
of the Johann Wolfgang Goethe University,
Frankfurt am Main, Germany

September 2020

For my beloved wife Aline

Requirements Engineering and Tool-Support for Security and Privacy

Dr. rer. nat. Sebastian Pape

Abstract

In order to address security and privacy problems in practice, it is very important to have a solid elicitation of requirements, before trying to address the problem. In this thesis, specific challenges of the areas of social engineering, security management and privacy enhancing technologies are analyzed:

Social Engineering An overview of existing tools usable for social engineering is provided and defenses against social engineering are analyzed. Serious games are proposed as a more pleasant way to raise employees' awareness and to train them.

Security Management Specific requirements for small and medium sized energy providers are analyzed and a set of tools to support them in assessing security risks and improving their security is proposed. Larger enterprises are supported by a method to collect security key performance indicators for different subsidiaries and with a risk assessment method for apps on mobile devices. Furthermore, a method to select a secure cloud provider – the currently most popular form of outsourcing – is provided.

Privacy Enhancing Technologies Relevant factors for the users' adoption of privacy enhancing technologies are identified and economic incentives and hindrances for companies are discussed. Privacy by design is applied to integrate privacy into the use cases e-commerce and internet of things.

Preface

When my research career started with a diploma thesis on the Ajtai-Dwork crypto system [144, 145], my focus was on the most theoretical aspects of cryptography. During my dissertation [150, 151] which addressed visual cryptography [149, 152] and non-transferable anonymous credentials [146–148], I broadened my view to also consider the environment in which the proposed approach would be applied. Naturally, the next step was to consider not only the environment itself, but also the economic and legal aspects, as well as the usability and user acceptability, treating privacy and security as cross-sectoral and interdisciplinary topics.

The habilitation thesis in front of you, "Requirements Engineering and Tool-Support for Security and Privacy", forms the basis of my journey through research on social engineering, security management and privacy-enhancing technologies. It has been written to fulfil the requirements for the academic title Dr. habil. (Doctor habilitatus) at the Faculty of Computer Science and Mathematics of the Johann Wolfgang Goethe University, Frankfurt am Main. The main part of this thesis was written in September 2020 although most of the research was conducted over a number of years. My habilitation thesis was submitted in September 2020 and accepted by the extended faculty council in January 2021 as a cumulative thesis. Depending on what the publishers allow, the related papers in the appendix could either be the final published versions or in some cases pre-prints, which would be accepted author versions.

I would like to express my deepest gratitude to Prof. Dr. Kai Rannenberg, who gave me the opportunity to pursue my habilitation at his chair and accompanied my academic career with helpful and supportive advice, a constant willingness for discussions, invaluable conversations and constant support. I would also like to take this opportunity to thank him most sincerely for writing an expert report on this habilitation thesis. I would like to extend my sincere thanks to the external reviewers Prof. Dr.-Ing. Felix Freiling and Prof. Dr. Melanie Volkamer for taking on and promptly delivering the expert reports. I also wish to thank everybody involved at the Faculty of Computer Science and Mathematics of the Johann Wolfgang Goethe University for their support and the smooth course of my habilitation process, in particular dean and head of the habilitation commission Prof. Dr.-Ing. Lars Hedrich as well as the commission's members Prof. Dr. Uwe Brinkschulte, Prof. Dr. Detlef Krömker, and Prof. Dr. Mirjam Minor.

Scientific work thrives on the suggestions, hints and criticisms of active and interested discussion partners. I would like to take this opportunity to thank all my colleagues at the Johann Wolfgang Goethe University and the projects I have been involved in for their wonderful cooperation. In particular, I am grateful to all my co-authors for the fruitful discussions and joint efforts: Dina Aladawy, Kristian Beckers, Sören Bleikertz, Maren Braun, Xinyuan Cai, Julian Dax, Trajce Dimkov, Veronika Fries, Ludger Goeke, Akos Grosz, David Harborth, Majid Hatamian, Vera Hazilov, Jan Jürjens, Dennis-Kenji Kipker, Jörg Lässig, Benedikt Ley, Fabio Massacci, Toni Mastelic, Federica Paci, Niklas Paul, Wolter Pieters, Volkmar Pipek, Alejandro Quintanar, Peter Schaab, Michael Schmid, Christopher Schmitz, Daniel Schosser, André Sekula, Jelena Stankovic, Mattea Stelter, Daniel Tasche, and Welderufael B. Tesfay.

To all the persons mentioned here, I would like to express my sincere and heartfelt thanks. Needless to say, any errors and inaccuracies are entirely my responsibility. I hope you enjoy reading this thesis.

Frankfurt, April 2021

Sebastian Pape

Contents

List of Figures

List of Tables

Chapter 1

Introduction

> It's important to recognize that you can't have 100 percent security and also then have 100 percent privacy and zero inconvenience.
>
> *Barack Obama*

With several data breaches and data leakages every month [99], data security and privacy issues have arrived in the middle of society. However, tackling security and privacy issues is not an easy task since both of them not only involve all technical layers, but are highly interdisciplinary too.

Often the boundary between security and privacy is blurred in both directions: If the used social-technical systems are not secure, all user data is at risk to leak if a breach occurs. In fact, the Cloud Security Alliance lists for its top threats to cloud computing data breaches as top threat of its last three reports [36, 37, 40]. Vice versa, if personal data leaks, that data can also be used for further attacks on the individuals or their companies, e. g. by attacks on the 'reset password' mechanism of many sites [108, 111, 175, 185] or other means of social engineering.

Experience has shown that security and privacy problems are often hard and even there is a solution in theory or academia, it is still not self-evident that the proposed solutions get to work in practice. On the one hand, there is a gap between research and practice [139] and even if solutions in academia exist they are often not applied in practice or only decades later. On the other hand, a proposed solution which is secure in theory, does not automatically imply that it is secure in practice. This holds for technical measures, e. g. cryptographic algorithms whose implementations can be attacked by side-channel attacks [115, 116], but also for humans which regularly struggle to use programs and mechanisms designed by engineers [69, 195, 212].

In order to address problems in practice, it is very important to have a solid elicitation of requirements, before trying to address the problem. In this thesis, specific challenges of the areas of social engineering, security management and privacy enhancing technologies are analyzed:

Social Engineering (cf. Sect. 2) The main challenge when counterfeiting social engineering is that all its defenses need to consider human behavior, which – contrary to technical systems – is in general not deterministic, but depends on a variety of other factors. While a variety of tools exists, most of them rather support attackers than defenders. This is not necessary the fault of the tools, since many of them were not designed for social engineering attacks. Besides the disadvantage on the tool side, it is also a hard task to raise awareness and train employees, since in general their main task is not fighting social engineering attacks off. Section Sect. 4 provides an overview of existing tools usable for social engineering and analyses defenses against social engineering. Additionally, serious games are proposed as a more pleasant way to make employees aware and to train them.

Security Management (cf. Sect. 3) One of the challenges for security management is that information security can only be measured indirectly [25], e. g. by using metrics and KPIs[1] which aim to approximate the real status of information security. Unfortunately, security management often goes together with compliance, which means that sometimes measures are not applied to increase the

security, but to demonstrate compliance in order to anticipate claims for damages should the company be successfully attacked. This also means that security risk assessment has to be a fundamental part of security management and often requires information security management systems to be implemented. Naturally, requirements differ for small and large companies. Sect. 3 analyses specific requirements for small and medium sized energy providers and proposes a set of tools to support them in assessing security risks and improving their security. Larger enterprises are supported by a method to collect security key performance indicators for different subsidiaries and with a risk assessment method for apps on mobile devices. Furthermore, as the currently most popular form of outsourcing, the selection of a secure cloud provider is discussed.

Privacy Enhancing Technologies (cf. Sect. 4) For privacy enhancing technologies the main challenge is their dissemination. Often companies do not want to integrate privacy enhancing technologies into their services, because their business model is built on collecting the users' data, or they think that they might need the collected data later, or because they simply do not know how to integrate them without harming usability or performance. On the other hand, even if stand-alone privacy enhancing technologies exist, the users' adoption is quite low and in particular for laymen it can be a cumbersome task to get them working [69]. Section Sect. 4 therefore aims to identify relevant factors for the users' adoption of privacy enhancing technologies. Besides that, economic incentives and hindrances are discusses and privacy by design is applied to integrate privacy into the use cases e-commerce and internet of things.

Section 5 elaborates on commonalities of the three areas and sketches future work. For a better overview, a mapping of papers to requirement elicitation and tool-support is provided in Tab. 1.1. However, there is no clear border since many of the papers on tools not only rely on previous results, but also include a short eliciting of requirements, e. g. an experiment.

Table 1.1: Mapping of Papers to Requirement Elicitation and Tool-Support

Section	Topic	Requirement Elicitation	Tool-Support
2	Social Engineering	A.3, A.4, A.5, A.10	A.1, A.2, A.6, A.7, A.8, A.9
3	Risk Assessment & Security Management	B.2, B.7, B.10	B.1, B.3, B.4, B.5, B.6, B.8, B.9
4	Privacy Enhancing Technologies	C.1, C.2, C.4, C.5, C.7, C.8, C.9, C.10	C.3, C.6

Chapter 2

Social Engineering

My work is a game, a very serious game.

M. C. Escher

The European Network and Information Security Agency, ENISA, defines social engineering as a technique that exploits human weaknesses and aims to manipulate people into breaking normal security procedures [143]. In most cases, maliciously motivated attackers aim to gain access to their victim's commercial, financial, sensitive or private information in order to use it against them or cause harm otherwise [8].

"The biggest threat to the security of a company is not a computer virus, an unpatched hole in a key program or a badly installed firewall. In fact, the biggest threat could be you [. . .] What I found personally to be true was that it's easier to manipulate people rather than technology [. . .] Most of the time organizations overlook that human element."[1] These words from Kevin Mitnick, a former hacker who now works as an IT security consultant, spoken in a BBC interview were made almost two decades ago and are still of utmost importance today.

The latest Data Breach Investigations Report [12] supports Mitnick's statement and reports another increase of financially motivated social engineering, where the attacker directly ask for some money, i. e. by impersonating CEOs or other high-level executives. Social engineering attacks represent a continuing threat to employees of organizations. With a wide availability of different tools and information sources [19], it is a challenging task to keep up to date of recent attacks on employees since new attacks are being developed and modifications of known attack scenarios are emerging. For example, during the last year, scammers have already varied their approach and also ask for purchase and transfer of online gift cards[2] in order to scam employees. Additionally, scammers also base attacks on the current news situation, such as COVID-19 Ransomware [182] or fake websites [15].

Furthermore, a social engineering attack is often only the first step of a larger attack, in which the attacker uses the information gained there for further attacks [12]. According to Milosevic [130], a social engineering attack itself consists of multiple phases as summarized in Table 2.1. In the first phase, the attacker conducts surveillance to identify persons with access to the information the attacker desires. The second phase focuses on finding out as much about these persons as possible to help the attacker to manipulate the victims. Based on that information, the attacker starts building a relationship to the victim (pretexting phase). Afterwards the attacker exploits the built up trust in the relationship and evaluates the gathered information in the post-exploitation phase.

However, most organizations have difficulties addressing this issue adequately. According to Mitnick, most companies rather purchase heavily standardized security products, such as firewalls or intrusion detection systems, than considering potential threats of social engineering attacks [132]. Peltier [166] supports this observation and states that technology-based countermeasures should be applied whenever possible. However, he also claims that no hardware or software is able to protect an organization fully against social engineering

[1] https://news.bbc.co.uk/2/hi/technology/2320121.stm
[2] https://twitter.com/sjmurdoch/status/1217449265112535040

Table 2.1: Overview of Social Engineering Phases by Milosevic [130]

Phase	Description
Pre-Engagement	Find targets with sufficient access to information/knowledge to perform an attack.
Intelligence Gathering	Gather information on each of the valid targets. Choose the ones to attack.
Pretexting	Use gathered information to build a relationship to the target. Gain victims' trust to access additional information.
Exploitation	Use the built up trust to get the desired information.
Post-Exploitation	Analyze the attack and the retrieved information. If necessary return to a previous phase to continue the chain of attack until the final information has been retrieved.

attacks. Furthermore, social engineering is highly interdisciplinary, however most defense strategies are advised by IT security experts who rather have a background in information systems than in psychology [184].

The remainder of this chapter is structured as follows:

- Sect. 2.1 discusses tools for and defenses against social engineering.
  - Sect. 2.1.1 describes a survey on tools for social engineering [19] (cf. Sect. A.5).
  - Sect. 2.1.2 surveys defense strategies and compares them to findings in social psychology [183, 184] (cf. Sect. A.3, A.4).
- Sect. 2.2 sketches the purpose and relations of the different serious games introduced in the subsections.
  - Sect. 2.2.1 describes the serious game HATCH, along with its two different applications [16, 17] (cf. Sect. A.1, A.2), a legal assessment of them [153] (cf. Sect. A.10), and a structured method to generate appropriate scenarios to adapt HATCH to different domains [94] (cf. Sect. A.8).
  - Sect. 2.2.2 describes the serious game PROTECT [72] (cf. Sect. A.7) and its predecessor Persuaded [7] (cf. Sect. A.6).
  - Sect. 2.2.3 describes the concept for a CyberSecurity Awareness Quiz [158] (cf. Sect. A.9).

The respective papers can be found in Appendix A and the author's contribution for each paper is indicated in Tab. **??** on page **??**.

2.1 Social Engineering Tools and Defenses

Even if companies are aware of social engineering attacks, they have only a limited number of tools available to support them. This might be one of the reasons for the aforementioned preference of heavily standardized security products. In Sect. 2.1.1, we will discuss the findings of a survey on tools for social engineering.

The alternative to hire penetration testing companies that *attack* the company's employees and clients in order to show weaknesses in their defenses does not seem to be a promising solution: Besides the need to address legal issues, which requires high effort upfront [211], experiments indicate that this approach might be counterproductive due to humans' demotivation when confronted with the testing results [53].

In order to provide an overview of further alternatives, besides trainings and awareness campaigns, we will compare defense mechanisms suggested in IT security and (social) psychology in Sect. 2.1.2. The idea behind the comparison is that social engineering tackles humans and while IT security is a rather new discipline, (social) psychology can be traced back to the ancient Greeks [196]. Therefore, we can expect to find further concepts to fight social engineering.

2.1.1 Survey on Tools for Social Engineering

In the process of social engineering described in Tab. 2.1, the first two phases are heavily based on information gathering. For that purpose, a number of tools are available. On the one hand, these tools may be used by a social engineer to prepare an attack. On the other hand, these tools could also provide an organization with an excellent alternative to pen testing or awareness trainings, as they allow to analyze possible vulnerabilities. For that purpose, we did a structured survey on the tools' capabilities [19] and contribute the following:

- a classification of existing tools regarding categories such as proposed purpose, price, perceived usability, visualization of results;
- a survey of information types retrieved by the tools regarding information about company employees and their communication channels, as well as related information e .g. company policies;
- a mapping of tools to certain types of social engineering attacks (phishing, baiting, impersonation).

For the mapping study, first for each information type (e. g. email, friends, (private) location, co-workers, company location) and each considered social engineering attack (phishing, baiting, impersonation) it was determined if it was of help for executing the attack. Then for each of the investigated tools (e. g. Cree-py, Maltego) or websites (e. g. LinkedIn, Xing) it was investigated which information they could provide. By combining these tables, finally Tab. 2.2 indicates which of the tools may be useful for which of the social engineering attacks.

Table 2.2: Tools vs. Attack Type Knowledge [19]

Information Type	Cree.py	Gitrob	KnowEm	LinkedIn	Maltego	Namechk	Recon-ng	Spokeo	theHarvester	Wayback Machine	Wireshark	Xing
Telephone Number							P					P
Friends			P,I	P,I	P,I	P,I						P,I
Personal Information	P,I		P,I	P,I		P,I		P,I				P,I
Private Locations	P,I							P,I				P,I
E-Mail				P	P		P		P			P
Instant Messenger			P		P	P		P				P
Co-Workers: New Employee				I	I							I
Co-Workers: Hierarchies				I			I					I
Lingo												
Facilities: Security-Measures		B,I									B,I	
Facilities: Company Location	B,I			B,I			B,I	B,I	B,I			B,I
Websites					P		P		P	P		

with P for Phishing, I for Impersonation, and B for Baiting

Taking a closer look at the table, one can notice that the only information type that social engineering tools do not provide today is the so-called *company lingo*, the abbreviations and specific words used in a company or domain. However, by analyzing postings in business oriented social network sites, it might be a matter of time when machine learning is applied and big data analysis will fill this gap.

Further results of our survey were, that none of the investigated tools or websites offered specific help for the defense against social engineering attacks. None of the tools provided any kind of risk assessment for the collected information on the employees themselves or on chief information (security) officers. Moreover, none of the tools was able to propose to remove certain information or to propose to add fake or bogus information to the publicly listed information. Fake information would allow to easily identify social engineering attacks later which might have relied on it.

Additionally, most of the tools were easy to use, opening the field not only to skilled professionals, but also to non-experts or even script-kiddies.

2.1.2 Mapping of Defenses

As we have seen in the previous section, most of the tools rather support the attacker than the defender. Therefore, it's worth to have a closer look at different defenses against social engineering.

For that purpose, we surveyed the state of the art [183, 184] from a computer science, namely IT security viewpoint, as well as from the viewpoint of social psychology. Following Kruger and Kearney [119], social engineering awareness was considered to consist of the three dimensions knowledge, attitude and behavior. Therefore, the identified defense mechanisms were mapped to this three dimensions. Furthermore, defense mechanisms from IT security and social psychology can be mapped against each other as shown in Tab. 2.3.

Table 2.3: Comparison of Defense Mechanism Suggested in IT Security and Social Psychology [184]

Dimension		IT Defense Mechanism	Psychological Defense Mechanism
Knowledge	Attitude	Policy Compliance	–
		Security Awareness Program	Forewarning
		–	Persuasion Knowledge
		–	Attitude Bolstering
		–	Reality Check
	Behavior	Audit	–
		–	Inoculation
		–	Decision Making

When comparing the different mechanisms, it is visible that defense mechanisms within IT security have not reached their full potential, yet. Within the attitude dimension, they mostly consist of security awareness trainings and programs and the definition of security policies. In comparison, social psychology offers with forewarning a mechanism similar to awareness raising. However, persuasion knowledge (including knowledge about persuasion strategies as well as counter tactics) and attitude bolstering (which relies on a good knowledge on security policies and its implication to create a bolstering mind-set) go far beyond the described defense mechanisms from IT security. Last but not least, carefully designed reality checks, could help people realizing that in fact they are vulnerable. However, reality checks have to be carried out very carefully in order to avoid frustration and create similar effects than those of social engineering penetration testing.

From a behavioral perspective, audits (including penetration testing specifically for social engineering) are a typical defense mechanism within the behavior dimension. However, since in general the penetration testers focus strongly on the detection of attacks and not on any kind of trainings or reality checkings of their victims, besides the aforementioned effort to set it up, care has to be taken not to demotivate employees. On the other hand, from a social psychological point of view, inoculation, e. g. putting employees in a similar situation a social engineer would put them in to train counter arguments, and training on decision making, i. e. avoiding impulsive decisions, which often benefit social engineering attackers, might let the employees persist a real attack.

Altogether, the gap analysis shows that defense mechanisms from IT security can be improved, and in particular consider the behavior dimension for to little.

Another contribution of the literature survey is a review of psychological principles which support impulsive decisions and avoid deeper reasoning. By mapping them to the applicability of (psychological) defense mechanisms, we find that attacks exploiting authority, social proof or distraction are mainly defendable through the dimension of attitude and attacks based on liking, similarity, deception, reciprocation and consistency require a training of both dimensions, attitude and behavior (cf. Tab. 2.4).

As a result, we recommend to integrate persuasion knowledge and resistance trainings into trainings or awareness campaigns fighting social engineering. Furthermore, attitude bolstering has been shown to be

effective in decreasing the effectiveness of persuasion attempts [218], and could be vital when users are not only shown their failures but also their successful attempts to prevent a social engineering attack.

Table 2.4: Mapping of Defense Mechanisms Against Attacks Based on Psychological Principles [184]

Dimension		Defense Mechanism	Authority	Social Proof	Liking, Similarity, Deception	Commitment, Reciprocation, Consistency	Distraction
Knowledge	Attitude	Persuasion Knowledge	✓	✓	✓	✓	✓
		Forewarning	✓	✓	✓	✓	✓
		Attitude Bolstering		✓	✓	✓	
		Reality Check			✓	✓	
	Behavior	Inoculation			✓	✓	
		Decision Making	✓	✓	✓	✓	✓

2.2 Serious Games on Social Engineering

In order to address the issues identified and developed in the previous sections, we have designed three serious games. Serious games have built a reputation for getting employees of companies involved in security activities in an enjoyable and sustainable way. While still preserving a playful character, serious games are used for e. g. security education and threat analysis [52, 197, 198, 214, 215]. Since at that time, none of the games was specifically developed for social engineering, all three games proposed in this section, aim to address social engineering, although in a different way (cf. Fig. 2.1). We start with a brief overview of the games and their relation in this section. They are described in more detailed in the following subsections.

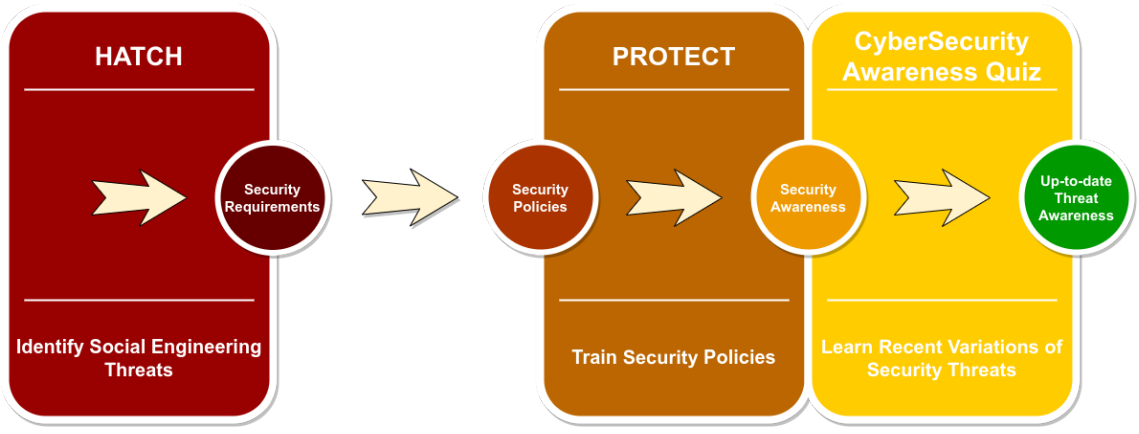


Figure 2.1: The Relation between HATCH [16], PROTECT [72] and CyberSecurity Awareness Quiz [158]

HATCH [16, 17] (cf. Sect. 2.2.1) aims to identify social engineering threats and develop them to security requirements. Since we noticed that most tools for social engineering do not help the defenders, the main idea of this game is to support the defenders in a systematic threat elicitation. Players develop attacks on their colleagues' based on their existing knowledge of work processes, skills and preferences. As a result, a list of possible SE threats is generated that can be used to improve work processes and security policies. The

advantage over a threat analysis by experts is that the employees of a department or a company know the real work processes very well, so it is easier to train them in social engineering than to have experts study all work processes. Furthermore, when asked about real processes, many employees will most likely not reveal what they are really doing but demonstrate their knowledge about how the process looks like in the official process definition.

While HATCH helps to develop and refine security policies, PROTECT [72] (cf. Sect. 2.2.2) aims to offer the employees an environment where they can learn and train the application of defenses. In the long run, the game raises the employees' security awareness and it also helps to bolster their attitudes.

Since PROTECT is based on security policies, it is naturally somewhat generic and can not address all recent variations of certain attacks. Therefore, the idea of the CyberSecurity Awareness Quiz [158] (cf. Sect. 2.2.3) is to have a quiz based on latest attacks and their variations. After a new attack or variation emerges, all it takes is the development of a new question which then can be used within the quiz immediately.

Besides the interplay of the three games itself, which already provides a tool chain to defend against social engineering, it is also important to allow the integration of the games into a more general training platform, such as the THREAT-ARREST [117] advanced training platform (cf. Fig. 2.2). The THREAT-ARREST[3] project received funding from the European Union's Horizon 2020 research and innovation program and aims to develop an advanced training platform incorporating emulation, simulation, serious gaming and visualization capabilities to adequately prepare stakeholders with different types of responsibility and levels of expertise in defending high-risk cyber systems and organizations to counter advanced, known and new cyber-attacks.

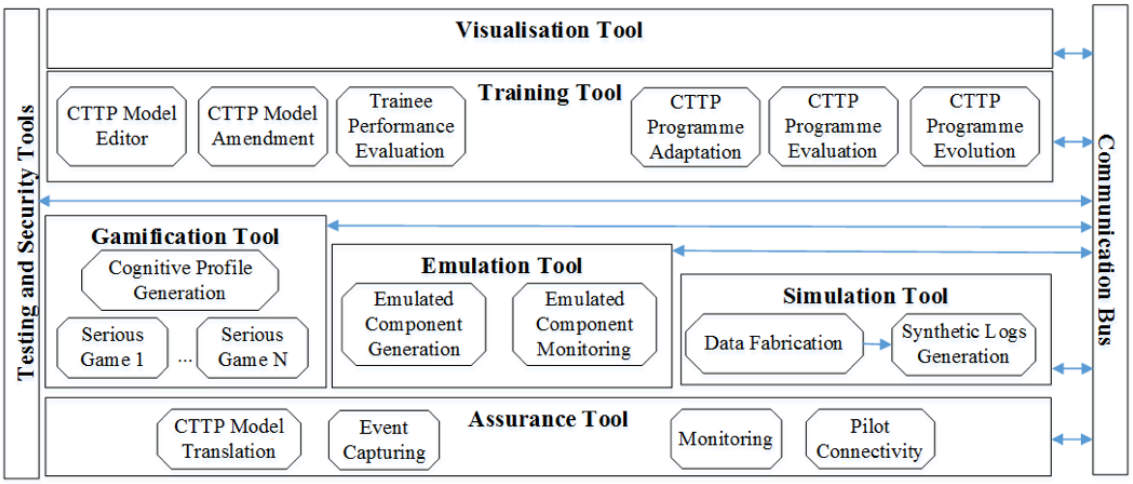


Figure 2.2: The THREAT-ARREST Advanced Training Platform [117]

The integration into the platform allows to combine the efforts within the serious game with other components such as the emulation and the simulation tool. This contributes to a continuous evaluation of the individual trainees' performance and the effectiveness of the training programs. Within the platform for each trainee results of the serious games, the emulation, the simulation and the training tool are brought together to spot possible gaps in the employee's knowledge or awareness. If knowledge gaps are identified, it can be checked if there already exists a training on the specific topic as serious game, simulation or emulation of the cyber range system. If no appropriate training can be identified, this might indicate the need of producing a new training, tailored to the organizational needs and the trainee types.

2.2.1 HATCH

Hack and Trick Capricious Humans (HATCH) is a physical (tabletop) serious game on social engineering [16, 17]. The game is available in two versions, a real life scenario and a generic version. Each version of the

[3] https://www.threat-arrest.eu/

game pursues a slightly different objective: The real life scenario is aiming to derive social engineering security requirements of a company or one of its departments. Therefore, a real environment is modeled and players attack their colleagues in order to identify real attack vectors. The generic version of the game aims to raise the players' awareness for social engineering threats and educate them on detecting this kind of attacks. In order to not unnecessarily expose and blame colleagues during a training session, it is based on a virtual scenario with personas as attack victims [153]. The initial scenario consists of a layout of a medium-sized office and ten employees as personas, printed on cards that contain fictional descriptions of them. By definition, personas are imaginary however, realistic descriptions of stakeholders or future users of a service or product, who have names, jobs, feelings, goals, certain needs and requirements [62].

In both versions two decks of cards are used: psychological principles and social engineering attacks. Psychological principle cards state and describe human behaviors or patterns that are often exploited by social engineers, as for example: 'Distraction - While you distract your victims by whatever retains their interests, you can do anything to them'. The psychological principle card patterns are based on the work of Stajano and Wilson [200], who describe why attacks on scam victims may succeed. We extended the set of behavioral patterns by patterns found in work on social engineering from Gulati [75] and Peltier [166]. On the other hand, the social engineering cards name and define some of the most common social engineering attacks, for example dumpster diving, which is 'the act of analyzing documents and other things in a garbage bin of an organization to reveal sensitive information'. The used attack techniques are mostly based on the work of Krombholz et al. [118]. Again, we extended the set of attack techniques by work of Gulati [75], Peltier [166], and Chitrey et al. [34].

When playing the game, each player draws one psychological principle card and three social engineering attack cards and reads the respective descriptions. Each player has then the task to choose a victim[4] which fits to the psychological principle card and to elaborate an attack by using one of the social engineering attack cards which matches victim and psychological principle best. Players take turns to reveal their cards and describe the social engineering attack they came up with. Other players discuss the proposed attack and award points for attack's feasibility and viability and rate if it is compliant with descriptions of this player's cards. The total score of each player is calculated by the end of the group rating and the player with the highest score wins the game. At the end of the game, all players briefly reflect on proposed social engineering attacks and derive potential security threats. The following list provides an overview of the steps of the game:

1. Each player draws a card from the deck of *human behavioral principles*, e. g. the "Need and Greed" principle.
2. Each player draws three cards from the deck of the *social engineering attack techniques*, e. g. phishing.
3. Each player develops an attack targeting one of the personas in the scenario based on the drawn cards.
4. Each player presents his/her attack to the group and the other members of the group discuss if the attack is feasible.
5. The players get points based on how viable their attack is and if the attack was compliant to the drawn cards. The player with the most points wins the game.
6. As debriefing, the perceived threats are discussed and the players reflect their attacks.

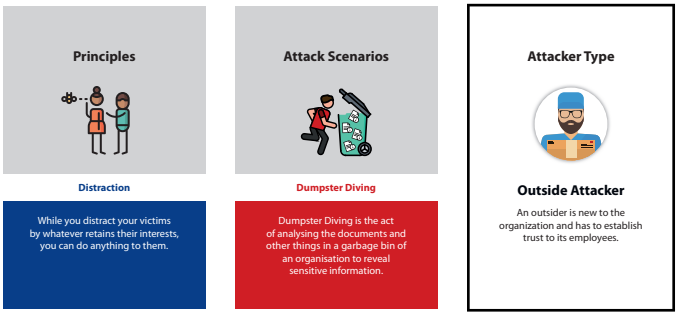
Figure 2.3 shows samples of the cards. The original card layout is shown in Fig. 2.3a and a version developed later with the help of Kristina Femmer, a professional designer, is shown in Fig. 2.3b. As discussed, the game can be played either with an imaginary (virtual) scenario or a (realistic) scenario that reflects the real working environment. We describe both scenario types in the following in more details.

Realistic Scenarios

The basic gameplay of HATCH has already been described above. For the realistic scenario, the aim was to identify a list of social engineering threats. Players should develop attacks on their colleagues based on their existing knowledge of work processes, skills and preferences. In order to foster creativity, a game plan

[4]depending on the version either a colleague or a persona

(a) Card Version 1 [16]

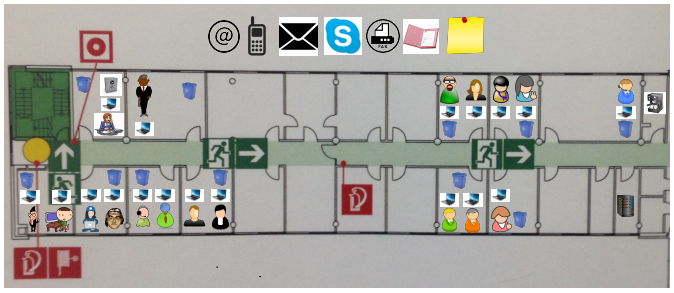


(b) Card Version 2

Figure 2.3: HATCH Cards:Psychological Principle, Social Engineering Attack, Attacker Type

is developed based on an existing emergency and escape plan (cf. Fig. 2.4a). Emergency and evacuation plans are a good source to build upon since they include a site plan which suits our needs and their layout is standardized [102]. Furthermore, they are publicly available in corridors, need to be updated frequently, and are – depending on the type of building – in most countries required by law (cf. [29, § 4 Abs. 4]). This way, a game plan can be created with low effort, e. g. by just adding images or icons of the co-workers and some assets (cf. Fig. 2.4b).

(a) Emergency and Escape Plan



(b) Adapted Game Plan

Figure 2.4: HATCH: Adaption of Emergency and Escape Plan for the Game

Besides training and awareness raising, the result is a list of possible social engineering threats that can be used to improve work processes and security policies. The advantage over a threat analysis by experts is that the employees of a department or a company know the real work processes very well, so it is easier to train them in social engineering than to have experts study all work processes. Beckers and Pape [16] showed that the realistic scenario was helpful for the elicitation of context-specific attacks by utilizing the domain knowledge of the players and their observations and knowledge about daily work and processes.

Virtual Scenarios

Virtual scenarios are used when HATCH is used for training and awareness purposes [17]. The basic gameplay of HATCH with a virtual scenario is the same as with a realistic scenario. However, instead of a plan of the real working environment along with co-workers, a map of a virtual environment is used. The virtual environment consists of a plan of a department or company (see Fig. 2.5) and for each of the employees shown in the plan there is a persona card that outlines the basic characteristics of the employee (see Fig. 2.6). The players' task now is to come up with an attack that is as plausible as possible on the basis of the drawn cards and that exploits the characteristics of the employees present in the game. The attack found is then evaluated for plausibility by the players.

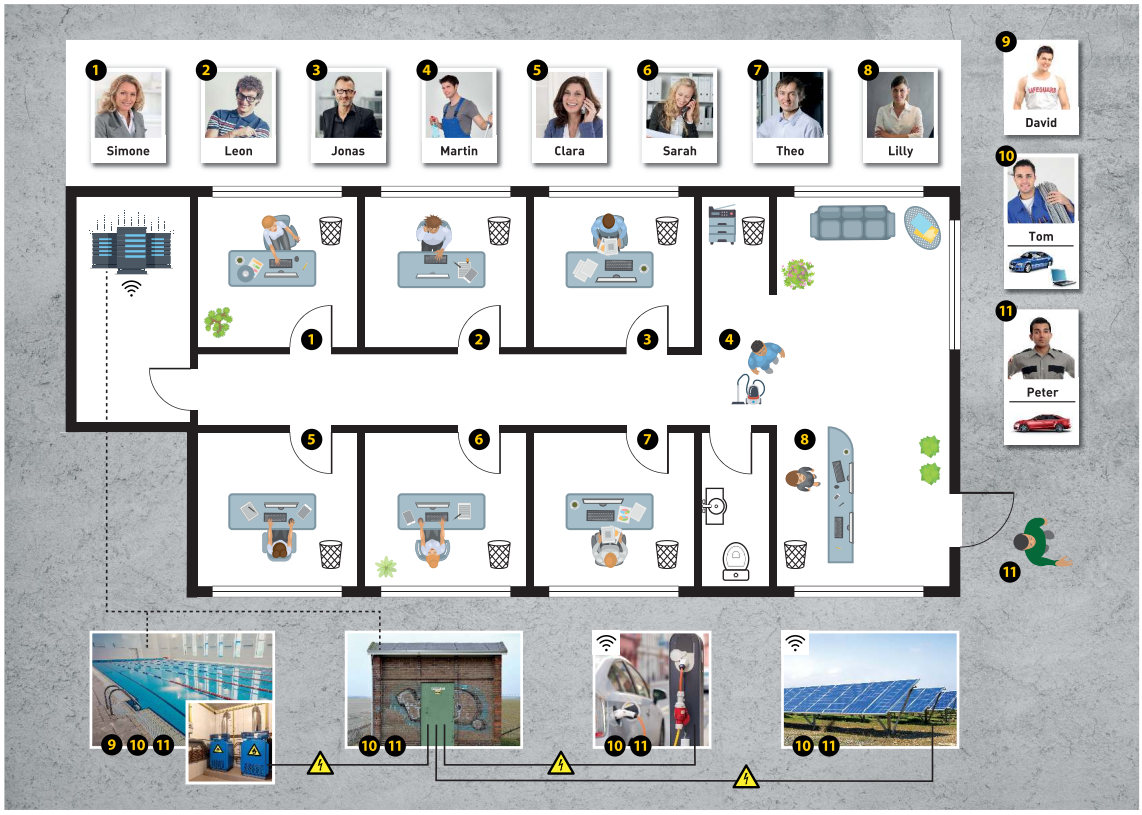


Figure 2.5: HATCH: Scenario for an Energy Provider

Besides the initial virtual scenario with a simple office environment, meanwhile scenarios for a maritime environment, an energy provider and a consulting scenario [94] exist. They all contain a basic map along with detailed persona descriptions on additional cards. Like the cards, the initial versions of the scenarios have been reworked with the help of a professional designer.

Legal Assessment

It is generally accepted that management has a legal obligation to maintain and operate IT security measures as part of the company's own compliance - this includes training employees with regard to social engineering attacks. Therefore, at a first glance, the use of a serious game for awareness raising and training against social engineering attacks seems to be fine. However, on the other hand the question is whether and how the employee must tolerate associated measures and, if necessary, also participate in them. The field of conflict between the employee's freedom and the company's security involves issues relating to labor law, data protection law, as well as for corporate compliance and corporate governance.

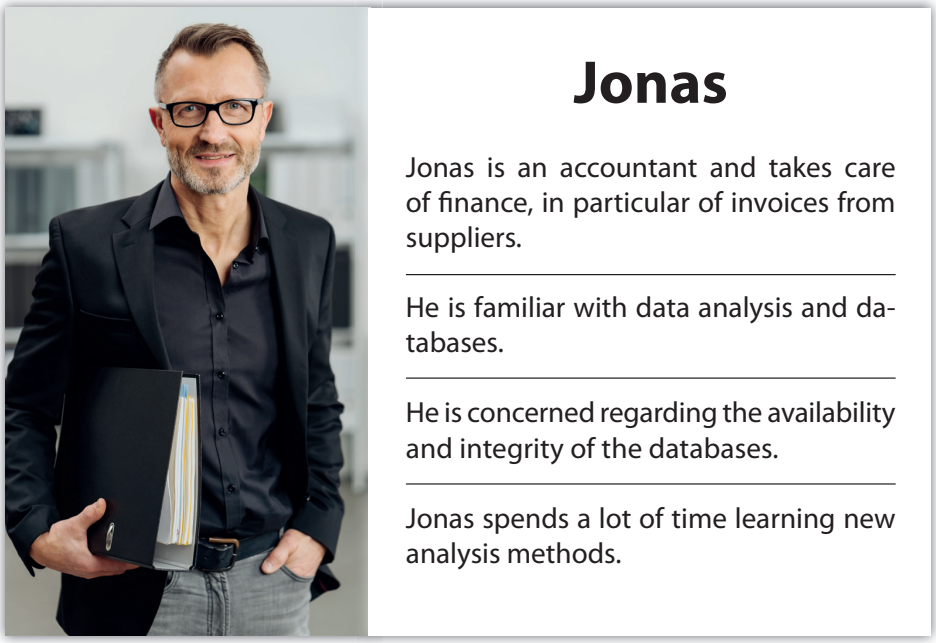


Figure 2.6: HATCH: Persona Card for Jonas, an Accountant

While there are reports on the use of serious games in the corporate sector [56], the body of literature specific to serious games aiming to raise awareness and allow security training is rather low. Regarding compliance and serious games, there is a lot of work, but only on using serious games to increase the compliance and not on the compliance of serious games. In the area of social engineering, most of the work is focused on social engineering penetration testing. Hatfield [93] discusses the ethics of social engineering penetration testing, and Kuhn and Willemsen [121] and Zimmer and Helle [220] discuss social engineering penetration testing from a legal perspective towards labor law. Therefore, we have investigated the legal challenges to make use of the game HATCH, and in particular the circumstances for HATCH's two different scenario types [153][5].

As a result, our legal assessment showed large differences in the assessment of the two different scenario types. In the realistic scenario, employee's personal characteristics are part of the game, thus care needs to be taken to not unnecessarily expose the personality of the employees, e. g. by accidentally revealing details of another employee such as long breaks, political, religious or sexual preferences not known to other players before – which could all be part of the game. Furthermore, it can not be ensured that some players do not use the environment of the game for some (additional) harassment at work. This even holds when the employees ask for or volunteer to play the scenario with a realistic environment, where they would suggest social engineering attacks on each other. Thus, for training and awareness raising, the virtual scenario should be used. On the other hand, if the employer can demonstrate a reasonable interest, i .e. if the game is used for threat analysis, the use of the game with a realistic scenario may be admissible.

Scenario Creation

Awareness campaigns benefit from addressing the target audience as specific as possible [11]. Transferred to HATCH, this refers mostly to the virtual scenarios and the need to have them as specific as possible. For that purpose, we investigated how to systematically developed a new scenario suitable for consulting companies. Our approach [94] also tackles the problem, that although many serious games for IT security exist, it is still hard to find an accurately fitting serious game for the environment of a specific organization.

In 2011, Faily and Flechais [62] introduced a method for developing personas that is based on grounded theory, a "[. . .] systematic, yet flexible guideline for collecting and analyzing qualitative data" [33]. Our proposed scenario creation process and consists out of 6 steps as shown in Fig. 2.7:

[5]based on previous work by the same authors [113]

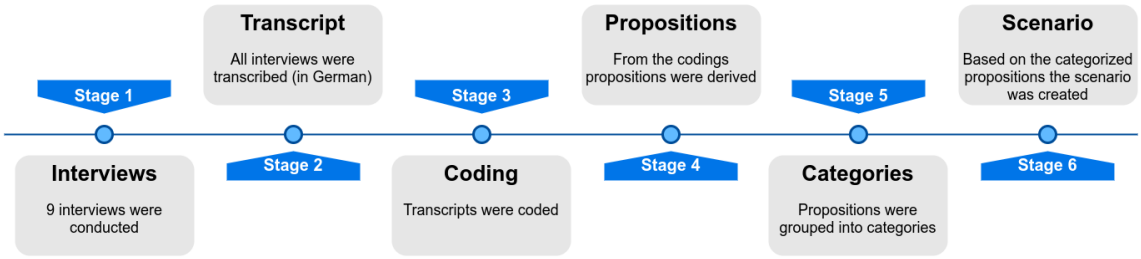


Figure 2.7: HATCH: Overview of Scenario Creation Process [94]

Conduct interviews with relevant stakeholders (stage 1) and transcribe them (stage 2). Code the answers in two rounds: open and axial coding (stage 3). Open and axial coding are typical for qualitative analyses: For open coding textual data is analyzed line-by-line to identify certain phenomena and attaching adequate codes to it. For axial coding previously assigned codes are examined to identify certain relationships among them and summarizing them into concepts and categories [42]. Following Faily and Flechais' method [62], develop propositions from codes (stage 4), such as 'more consultants are hired for project than clients' and 'generally, the consulting team consists of 4 to 5 people'. Summarize these propositions, assign them to concepts and categorize them (stage 5). As the last step, select appropriate propositions and use them as characteristics for persona narratives (stage 6).

The result of the evaluation of our method for creating a new scenario for HATCH was that it was effective: All participants of the evaluation sessions agreed that the derived scenario and its personas are realistic. However, since it was also very time-consuming, the required effort only makes sense if the scenario can be used several times by an organization or can be transferred to another, similarly structured organization.

2.2.2 PROTECT

The serious game PROTECT [72] builds on its predecessor Persuaded [7], thus both games share the same gaming principle. Players draw cards in a patience like manner from a pile and besides special cards, the pile contains attacks and defenses. If an attack is drawn, the player gets confronted with a possible social engineering threat and has to select a defense mechanism. The correct defense mechanism is a pattern of behavior ensuring a secure outcome, e. g. as described in a security policy. An example of the user interface and for a presented attack is shown in Fig. 2.8.

In order to make the game more challenging and increase the long-term motivation to play the game, in the basis version, players need to ensure that they have the correct defense card on their hands when an attack is drawn from the pile of cards. For that purpose, they may use special cards to view the next three cards on the pile or to discard the next card on the pile. By playing anticipatorily, players can navigate through the pile, collecting defense cards and discarding attacks they do not have a correct defense card, yet.

PROTECT is an advancement of Persuaded in several directions. First of all, Persuaded was more or less static in the way that the cards were represented by images, making it quite difficult to build a new deck. PROTECT allows to define card decks in its configuration file, and therefore can be adapted with low effort to new scenarios or security policies. As a consequence, several different card decks exist, e. g. scenarios for maritime transport or electronic cancer registration domains. Furthermore, the game play can be changed by configuring various game settings to allow a progression between difficulty levels and various other challenges to allow the players to get familiar with attacks and defenses, but also keep the players' long-term motivation to start new games up. This makes PROTECT a family of games, with Persuaded being a specific member of the game family.

As a further enhancement, all configuration options are accessible via an application programming interface, allowing PROTECT not only to serve as a stand-alone application, but being easily embedded into a training platform. The training platform, e. g. the THREAT-ARREST platform, can then control the game's difficulty by changing configuration parameters based on the players achievements in previous games or in other trainings within the platform.

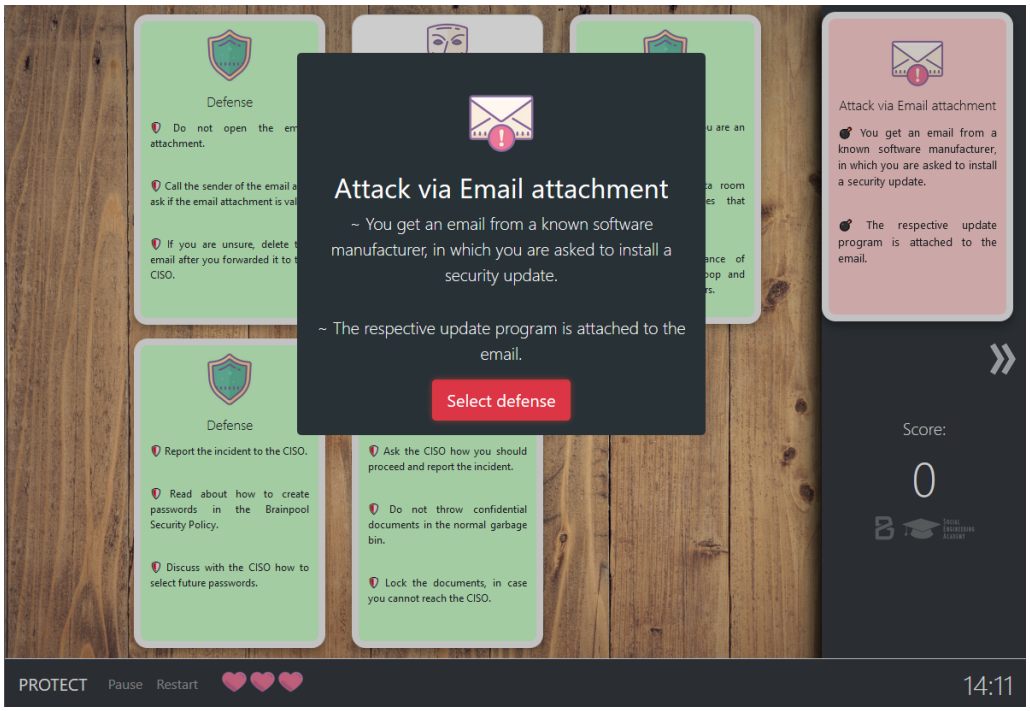


Figure 2.8: PROTECT [72]: Graphical User Interface

We have evaluated PROTECT by asking five practitioners to play the game and provide us feedback. The feedback was in general good, in particular emphasizing that after players are familiar with the game, the game contributes to bolstering the players' attitude by making them confident that they might be able to defend against certain social engineering attacks in future. Although, the game is of course less complex than reality, it also contributed to inoculation by letting the players react repeatedly on a limited number of attacks.

2.2.3 CyberSecurity Awareness Quiz

A general challenge for serious games is to cope with new attacks and variations of attacks. For most of the existing games, including PROTECT, it is lots of effort to adapt the game in a timely manner. Therefore, the idea of the CyberSecurity Awareness Quiz [158] is in particular to allow fast update of the game's content to cover recent threats.

During the game conceptualization, we defined the following requirements: i) As discussed beforehand, the game should refer to recent real-world threats. ii) Since we expect only a reasonable amount of new attacks, the game should be lightweight, short and playable on mobile phones with the idea that it could be played occasionally (e. g. when traveling in trams or subways). As a result, we decided to aim for a quiz-like game as shown in Fig. 2.9. Since the game type is straight forward, players answer questions and can either play single-player to compete for a high-score or have several multi-player modes to compete against each other, the main focus in this section is on a systematic process to create questions based on current affairs and attacks observed in the wild.

The first step of the process consists of the procurement of information with respect to current social engineering attacks as shown in Fig. 2.10. Within that step, relevant sources, regularly publishing content related to social engineering attacks like news websites, websites about information security, websites of institutions, blogs or even twitter are collected, preferably in a structured manner (e. g. standardized formats like RSS[6]). These sites are then automatically checked with appropriate tools for updates on new attacks. As of now, the updated information has to be manually reviewed by a game content editor to assess if the content to the web feed is relevant.

[6]depending on the version RSS means: RDF Site Summary or Really Simple Syndication

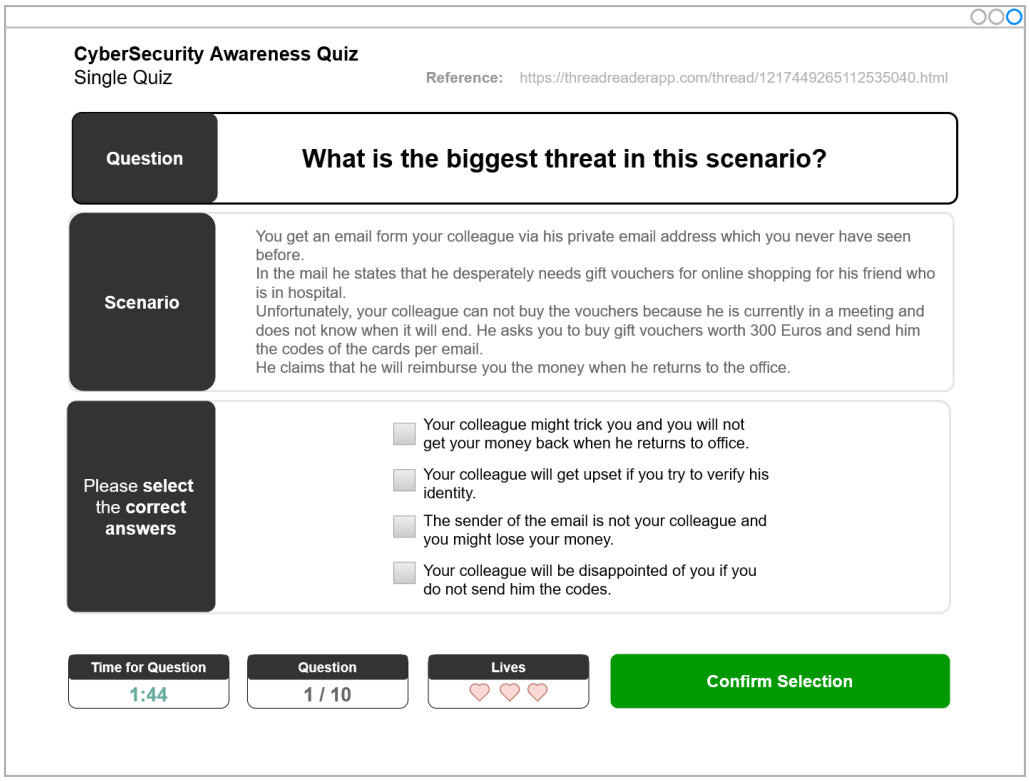


Figure 2.9: CyberSecurity Awareness Quiz [158]: Graphical User Interface

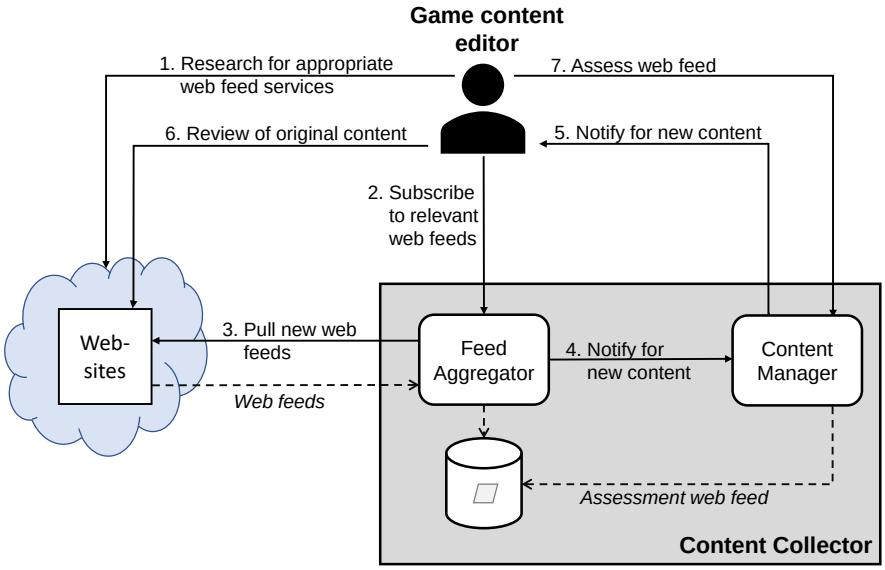


Figure 2.10: CyberSecurity Awareness Quiz [158]: Gathering and Analyzing Content about Attacks

If a relevant attack is identified, a question is formulated based on the attack. In order to allow the creation of quizzes based on a certain topic or on attacks popular in a certain time frame or region, metadata like the category of an attack (e. g. phishing) is assigned to the question. In the next steps, correct and incorrect answers are assigned to the question.

Several tools are provided to allow the quiz manager either to manually create quizzes or to create them based on certain metadata. Players can then choose out of a list of provided quizzes which topic they are

interested in. Alternatively, the CyberSecurity Awareness Quiz can be connected to a training platform, in the same manner as described for PROTECT in the previous section. This way, the platform is able to choose a quiz based on the player's performance in other parts of the training.

Chapter 3

Security Management

> There are risks and costs to a program of action – but they are far less than the long range cost of comfortable inaction.
>
> *John F. Kennedy*

Security risk assessments should be at the core of any digitally evolved organization. Often they are part of the organizations' constant effort for compliance. While a certification following ISO/IEC 27001 [104] in some domains, i. e. critical infrastructure, is mandatory, there are also economic reasons even for small and medium enterprises to get ISO/IEC 27001 certified [96].

Naturally, approaches and challenges will be different for small and large enterprises: Small enterprises often struggle with the necessary know-how since they can not afford entrusting someone full time with security. Therefore, already the introduction of an information security management system, which is required by the ISO/IEC 27001 standard, is a serious challenge for them. As a consequence, they often ask specialists for consultation. On the other side of the spectrum, large enterprises often have dedicated security specialists, but struggle more often with being split in several units. Their challenge is to setup a consistent security level across all units and allow their management to overlook the system and its security needs as a whole. For that purpose the collection of key performance indicators is vital for large enterprises.

However, there are also challenges regarding security which are common to small, medium and large enterprises. The outsourcing of processes and services has been part of strategic decisions since decades [172] and in particular the outsourcing of IT has been a trend in the 1990s [59]. Even with the occasionally reverse trend of backsourcing [213], this trend has been increased by cloud computing, which has been emerging as the new computing paradigm in the last ten years. Cloud Computing enables consumers to purchase on-demand, conveniently and cost efficiently computing power and storage capacity that can be rapidly provisioned and released with minimal management effort [138] from specialized providers. Recent studies claim that cloud computing has left the hype phase behind and can already be considered the norm for IT [27]. As cloud adoption is still a kind of IT outsourcing, it also comes with security concerns from the customers. On the other hand, in certain scenarios there might be also benefits to security since cloud service providers (CSPs) enjoys economies of scale in terms of security as well. Therefore, they are able to hire security specialists and thereby achieve a higher security level than most client companies would with an in-house data center [77, 109]. In either case, it is a challenging task for the customers to assess the security of cloud service providers and select the most secure one.

The remainder of this chapter is structured as follows:

- Sect. 3.1 focuses on the risk assessment for small and medium enterprises, in particular energy providers.
  - Sect. 3.1.1 focuses on reporting on the background and eliciting requirements for a tool supporting them with information security [46, 162] (cf. Sect. B.2, B.10).

  - Sect. 3.1.2 describes certain aspects of the developed tool, namely an inter-organizational security platform [189] (cf. Sect. B.3) and a lightweight risk assessment tool [188] (cf. Sect. B.9).

- Sect. 3.2 focuses on risk assessment for large enterprises
  - Sect. 3.2.1 describes an approach to compare the security levels of subsidiaries of large enterprises [187] (cf. Sect. B.4) with a refinement about different aggregation functions for security maturity levels defined for multiple assets [186] (cf. Sect. B.5).
  - Sect. 3.2.2 proposes an approach for the risk assessment of mobile apps [92] (cf. Sect. B.6).
- Sect. 3.3 discusses security assessments for customers of cloud service providers.
  - Sect. 3.3.1 first sketches the best practice in companies for security cloud service provider selection [155] (cf. Sect. B.7) to motivate the proposal of an semi-automated approach for secure cloud service provider selection [161] (cf. Sect. B.8).
  - Sect. 3.3.2 proposes a model for security assessments of cloud service providers [24] (cf. Sect. B.1).

The respective papers can be found in Appendix B and the author's contribution for each paper is indicated in Tab. **??** on page **??**.

3.1 Security Risk Assessment and Security Management for Small and Medium Energy Providers

Critical infrastructures are of vital importance to a nation's society and economy because their failure would result in sustained supply shortages causing a significant disruption of public safety and security. In 2016, malicious software in nuclear power plants was reported [201] followed by further reports, e. g. warnings about hackers attacking German energy providers in 2018 [194].

With the *European Program for Critical Infrastructure Protection* (EPCIP) and its counterpart, the German critical infrastructure protection program KRITIS [112] governments aimed to provide the ground for more secure critical infrastructures. The new regulation challenged critical infrastructure providers in many ways. Besides general challenges such as understanding the definitions and requirements (cf. [28, p. 150ff]), and challenges from other areas, i. e. coping with the energy transition, energy providers needed to register a contact point, establish processes to report security incidents, implement security requirements following a security catalog (§11 Abs. 1a respectively 1b EnWG), and establish and certify an information security management system (ISMS).

The SIDATE project [49] aimed to support small and medium energy providers to cope with the security requirements. Since most of the small and medium sized German energy providers were in a similar situation and they were not directly competing against each other, the idea was to support their collaboration using a web-based platform. For that purpose, we conducted a survey among all German energy providers and elicited criteria from energy providers on how such a platform should be designed [46].

3.1.1 Requirement Elicitation

Due to the new regulations in Germany energy providers are required to obtain IT security certificates. Especially small and medium-sized energy providers struggle to fulfill these new requirements. To get a general idea how they could be supported, we conducted a survey among all energy providers and had a series of workshops to discuss their needs and how a tool supporting needs to be designed.

Survey on Establishment of an Information Security Management System

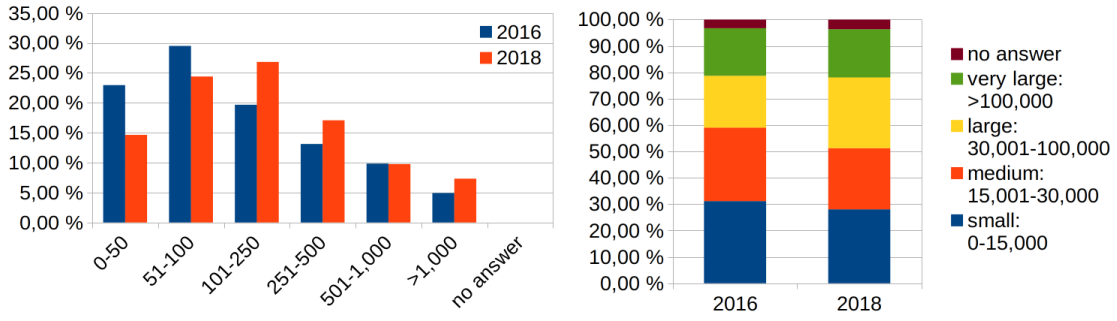
The investigation focused on the introduction of an information security management system (ISMS) and how German energy providers deal with information security in general. For that purpose, we surveyed German energy providers in autumn 2016 when they had just learned about the requirements and in autumn

2018, two years later and roughly half a year after they had to provide the certification of their ISMS. The new regulation offered the chance to have a closer look at a large amount of energy providers introducing an ISMS to get ready for certification at the same time [162].

The questionnaires covered sections about general information, organizational aspects, ISMS and ISMS maintenance (only in 2018), the office IT, and networking and organizational aspects about the industrial control system of the energy providers, which are reported in more details in technical reports [47, 48, 50, 156].

In 2016 (2018), we (physically) mailed to all 881 (890) energy providers listed in August 2016 (September 2018) [31] by the Federal Network Agency (German: Bundesnetzagentur or BNetzA), the German regulatory office for electricity, gas, telecommunications, post and railway markets [30]. We received a total of 61 (84) replies resulting in a response rate of 6.9% (9,4%).

We asked the energy providers about the number of supply points and the number of employees as shown in Fig. 3.1 to get an idea about their size. We checked with Spearman's rank correlation for similarities between the number of employees and the number of supply points and found for 2016 (2018) ρ-values of 0.725 (0.496) with p-values lower than 10^{-5} indicating a strong (moderate) relationship. Therefore, we argue that it is sufficient to consider the number of supply points and refer in the following to the size of an energy provider following the definition above. A comparison with the study from Müller et al. [135] shows that we had more small energy providers than they had.



(a) Number of Employees

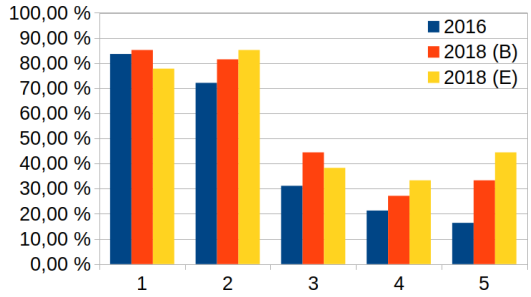
(b) Number of Supply Points

Figure 3.1: Size of the participating energy providers [162]

We also tested the similarity of the data for 2016 and 2018 with a two-one-sided t-test (TOST) [191] for the energy provider's size and since for $\epsilon = 0.5$ the p-value of 0.027 was within the 95% confidence interval, we assume that the participating energy providers were similarly distributed within both surveys.

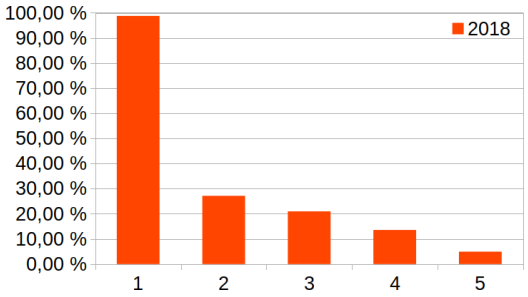
We also asked about the reasons the energy provider were implementing an ISMS and in 2018 additionally about the perceived benefits and the future expectations regarding the ISMS (cf. Sect. 3.2). Unsurprisingly, legal requirements were the largest factor. However, it also showed that most of them in 2016 were also expecting a security improvement, which even more of them reported as a benefit in 2018. The result is an indication that the German critical infrastructure protection program at least succeeded in making the energy providers implementing an ISMS. Most likely, most of the energy providers would not have implemented without being forced.

In order to get an idea about the status quo of the implementation of ISMS, we asked for each of the 18 phases if the phase was finished, begun, planned or not yet planned (cf. Fig. 3.3). Given the regulation, it came again not as a surprise that almost all energy providers had started in 2016 and most of them were finished in 2018. However, we were also aware that some of the small energy providers spend quite some effort in demonstrating that they do not fulfill the definition of a critical infrastructure, and they therefore do neither need to implement an ISMS nor get a corresponding certificate. This in line with Müller et al. [135] and one more time confirms that legal obligations were the main driver to implement an ISMS. Further results showed that the energy providers as a whole overestimated the needed duration for the implementation by roughly 20%. Furthermore, most of them only planned external support for the implementation of the ISMS but not for running it [156].



1 Fulfilling Legal Requirements
2 Improving Information Security
3 Better Representation of IT Processes
4 Better External Representation of IT Processes
5 (Re-)Structuring of Relevant Business Processes

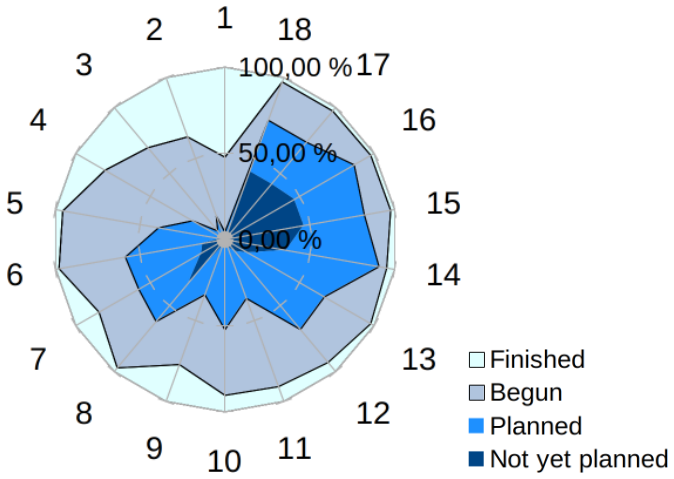
(a) Top 5 Reasons 2016 + Benefits and Expectations 2018



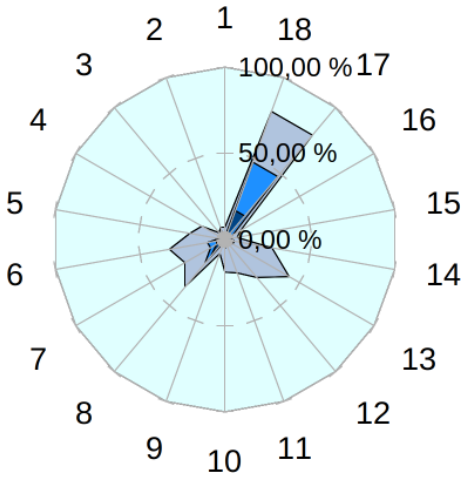
1 Legal Requirements
2 Business Processes are Depending on IT
3 Increased Threats
4 Public Discussion on IT-Security
5 Outsourcing of Services

(b) Top 5 Reasons 2018

Figure 3.2: Motivation, Benefits and Expectations to Implement an ISMS [162]



(a) 2016



(b) 2018

1 Target Setting and Scoping
2 ISMS Policy Development
3 Overview of the existing security architecture
4 Performing Risk Analyses
5 Elaboration of Catalog of Security Measures
6 Design of the New Security Architecture
7 Description of Quality and Risk Manag. Interf.
8 Development of a Migration Process
9 Elaboration of the Req. Documentation
10 Structure of the Security Organization
11 Implementation of Management Processes
12 Formulation of Security Architecture (Rules)
13 Measures of Sensitization and Training
14 Implementation of Security Measures
15 Final Project Scope Analysis
16 Preparation for Certification Auditing
17 Execution of Business and Organizational Audits
18 Incident-Management Support

Figure 3.3: Status of each ISMS implementation phase [162]

Requirement Elicitation for Inter-Organizational Security Platform

Besides the surveys, we also got some insights by workshops within the SIDATE project [49] with personnel from energy providers responsible for IT security [46, 189]. Since most of the German energy providers were in the same situation and they were not directly competing against each other, the idea was to support their collaboration using a web-based platform. For that purpose, we elicited criteria from energy providers on how such a platform should be designed [46] in the workshops.

We conducted three two-hour workshops with different stakeholder groups with in total eleven experts from eight energy providers. Most participants were IT security officers or IT managers from energy providers, but also representatives from national interest groups were present.

In the first workshop, we elicited the platform's requirements and the experts' expectations in a moderated discussion. The following modules were considered helpful by the experts: a wiki, a forum, a questions and answers module, a glossary, training modules for further education for security officers and other employees, checklists, a place to exchange documents, benchmarks, security assessment modules and a general module to support the launch of an ISMS.

Because the platform processes highly sensitive data, data privacy requirements had a very high priority for the stakeholders, e. g. having different user interface views to anonymize individuals and organizations to external experts, and having a restricted and moderated access for new members. Furthermore, the integration into existing workflows played a central role, e. g. the self-assessment should provide individual checklists and tools according to ISO/IEC 27001 [104] and should contribute to the internal information security audit. Besides that, the general usability of the platform was mentioned as essential requirement.

Based on these results from the first workshop, a design workshop with eight members from the project partners was conducted. As a result, the most relevant modules for the energy providers were identified and several mock-ups visualizing the platform's functionalities were sketched (cf. Fig. 3.4):

- A security assessment module, allowing the energy providers to assess and benchmark their security level.
- A security measures module, providing information and recommendation to energy providers (including the practical experiences by other energy providers) about security measures they can implement in order to strengthen their IT-security.
- A question and answer module, allowing the energy providers to share their experiences with both other energy providers as well as with external experts.

In the third workshop, the developed mock-ups were presented and the experts were asked for mandatory and nice-to-have requirements the platform had to fulfill to be usable for them. We clustered the answers into four major categories: (1) platform members and confidentially/data privacy, (2) integration into existing workflows, (3) general usability of the platform.

Platform Participants and Data Privacy As already discussed during the first workshop, participants had essential concerns about the privacy in respect to sensitive IT-security related data they would share across the platform. Interestingly, these concerns did not refer much to the platform itself or its operator but to other members. Participants were most worried about the participation of external experts like information security consultants or lawyers. Even if they saw an advantage in the qualified and skilled feedback from such persons, they were afraid of misuse of the platform for advertising purposes and non-reliable members could use the content of individual energy providers to identify them and exploit possible security flaws. This resulted in the following list of requirements:

R.1 The platform should support restricted and moderated access for new members, members need to be validated by the platform operator and have to agree to suitable terms of use in order to get access.

R.2 Some participants of the platform should not be able to see the corresponding author of content within the platform, e. g. external experts should not be able to identify energy providers. A reputation system should allow the energy providers to assess the quality of a contribution or the reputation of an external expert.

R.3 Alternatively, no third parties such as external experts should be allowed on the platform, but energy providers should be able to mark content as 'expert approved' to allow the indirect passing of experts' assessments and opinions.

Integration into Existing Workflows The effort necessary for using the platform should not exceed the potential benefit and be integrated into users' existing workflows. This resulted in the following list of requirements:

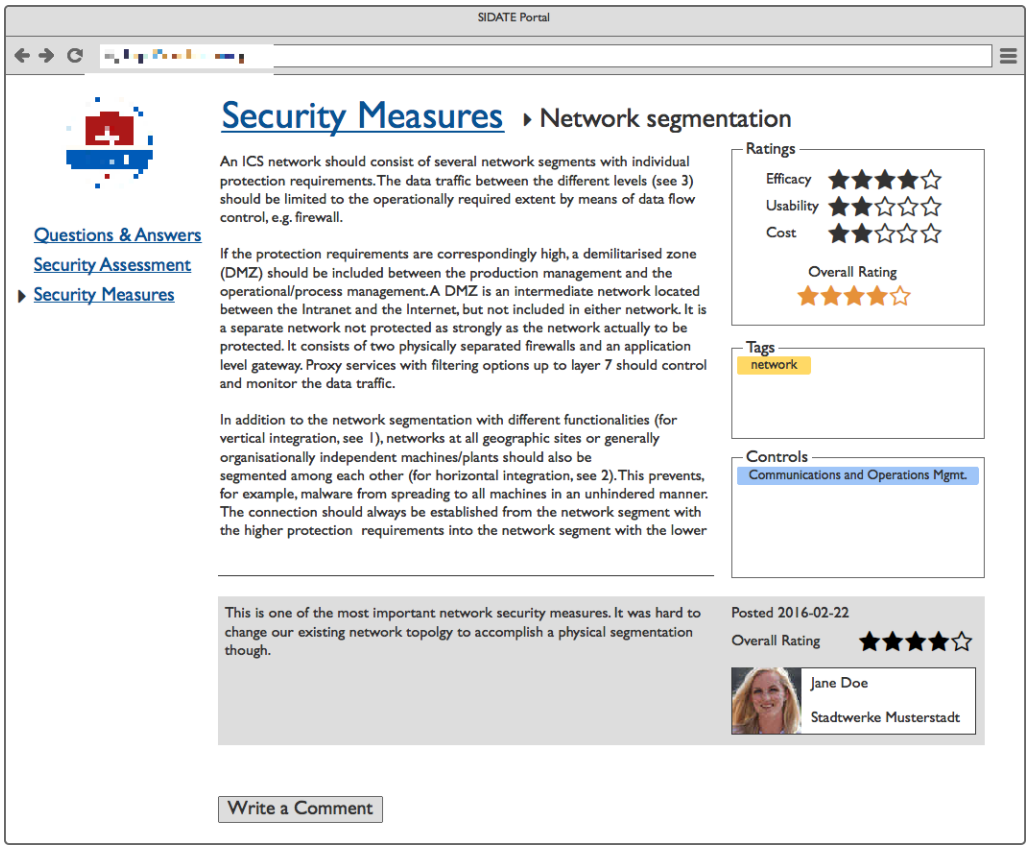


Figure 3.4: Portal Mock-up: Security Measures Module [46]

I.1 The self-assessment module should provide individual checklists and tools that help the users to implement required information security measures, i. e. fulfillment of statutory provisions such as the implementation of an ISMS according to ISO/IEC 27001 [104].

I.2 The self-assessment should contribute to internal information security audits, e .g. the regular validation of measures and processes.

I.3 The export from results of the self-assessments, e. g. for internal reports or other processes should be possible.

General Usability of the Platform The requirements focused on the usability were very general and not specific to the platform, such as expectations that the platform should be well-structured and maintained, and should have a moderator who ensures that new topics/questions are created in the right section and prevents duplicates.

3.1.2 Tool-Support

Based on the requirements elicitation, two different tools were developed: An inter-organizational security platform [189] and the risk assessment tool LiSRA [188]. However, LiSRA was connected to the platform via a REST API. This way LiSRA can make use of the data within the platform and users of the platform do not need another tool or website to make use of it.

Inter-Organizational Security Platform

In the previous section, we elicited requirements for an inter-organizational security platform with the idea that energy providers can exchange expertise when working on similar problems [51, 192]. We do not

discuss the requirements regarding platform participants and data privacy any further, since these were either organizational processes outside the platform or can be covered by the underlying framework Liferay [1].

The requirements regarding the integration into existing workflows were fulfilled in the following way. The self-assessment was based on maturity levels for security controls in ISO/IEC 27001 (I.1) as shown in Fig. 3.5, but can be further improved by a more specific standard (e. g. ISO/IEC 27019 [105] for the electric sector). By referring to the security maturity levels of the COBIT framework [101], which are also

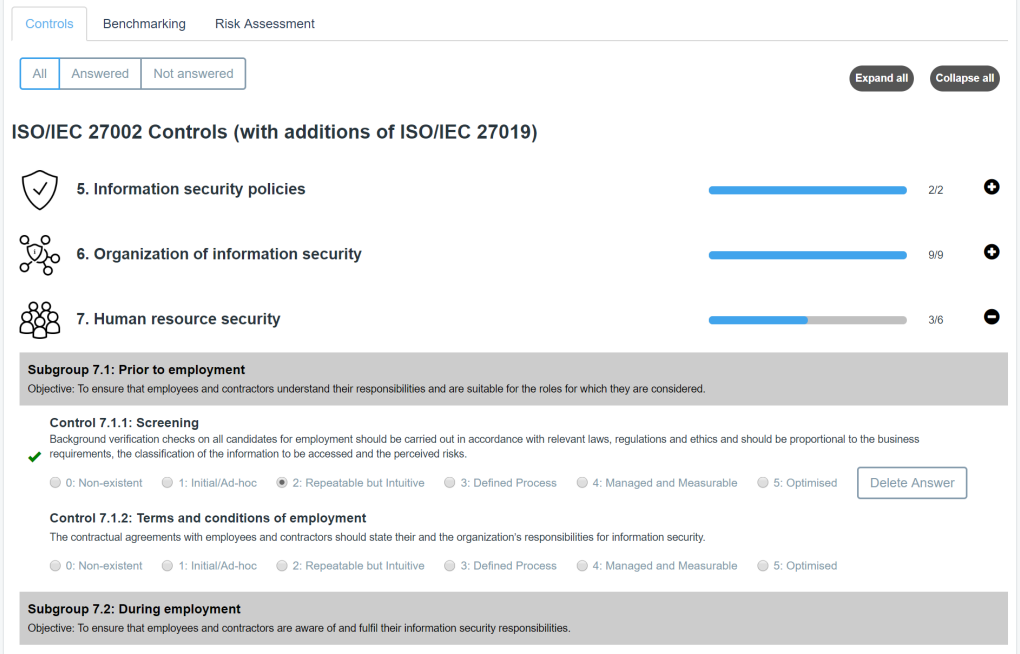


Figure 3.5: Portal: Input Section [189]

defined in ISO/IEC 15504 [103], the self-assessment can also contribute to internal security audits (I.2). Individual checklists and tools were covered by a document exchange module (I.1) which allows the member to exchange and discuss documents. In order to keep the information up to date, a portlet was developed, which asked users for missing or outdated maturity levels as shown in Fig. 3.6. The export function of the security maturity levels was implemented via a REST API (I.3).

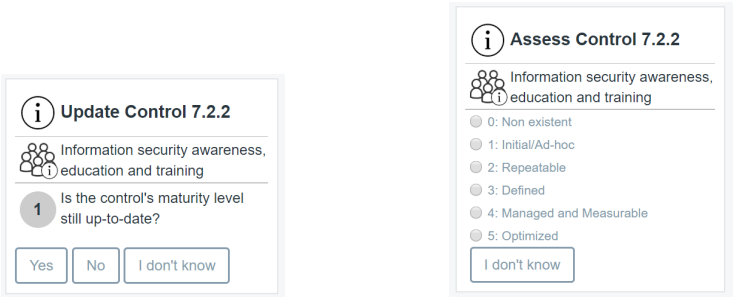


Figure 3.6: Portal: Modules for Updates to Maturity Levels [189]

Figure 3.7 shows the benchmark functionality. Besides an overview for each of the ISO/IEC 27001's control (sub)groups, it shows the maturity levels for other users of the platform along with scale where the user's own maturity level is compared to the others. Note that the recent security maturity level can also

[1] https://www.liferay.com/

easily be changed here, without the need to change to the input section. In contrast to the risk assessment tool described in the next section, there is no further evaluation besides the comparison to other users here.

Figure 3.7: Portal: Benchmarking Section [189]

Risk Assessment Tool: LiSRA

As already discussed in the previous section, the lightweight security risk assessment (LiSRA) [188], is integrated via a REST API into the inter-organizational security platform as shown in Fig. 3.8. The bar on the top shows the calculated risk value in the range from 0 to 1. LiSRA is based on attack trees, and therefore on the lower half of the user interface each attack scenario can be examined in form of a risk value per scenario. The corresponding attack trees can also be investigated in detail. On the lower right is a list of controls answered by the user and relevant for the investigated attack tree.

LiSRA is designed with a particular focus on the special needs for SMEs. Therefore, a key requirement is to mainly use already existing data and to keep the user's input to a minimum but to ensure good analysis results at the same time. To meet the requirements, LiSRA expects input from both users and domain experts who may be associated with the platform provider. The general concept consist out of four phases and is illustrated in Fig. 3.9:

Phase 1: Expert Input LiSRA assumes that organizations within a particular domain are basically exposed to similar attacks, e. g. the National Electric Sector Cybersecurity Organization Resource (NESCOR) [137] lists domain-specific attack scenarios for the electric sector. Therefore, in the first phase domain experts initially set up the framework for particular domains (e. g. the electric sector) by constructing parameterized attack trees that are linked to security controls. In a later step the user can select the domain in which his organization operates so that the risk assessment only considers attack trees that are relevant for the respective domain.

Phase 2: User Input The only user inputs required are the maturity levels of the organization's security controls. They are used to model the implemented security practices of the organization in a lightweight manner. For many organizations this only causes little extra effort because they have already collected these information, e. g. within their ISMS. As already mentioned, LiSRA is integrated into the

Figure 3.8: Portal: Risk Assessment Section by LiSRA [188]

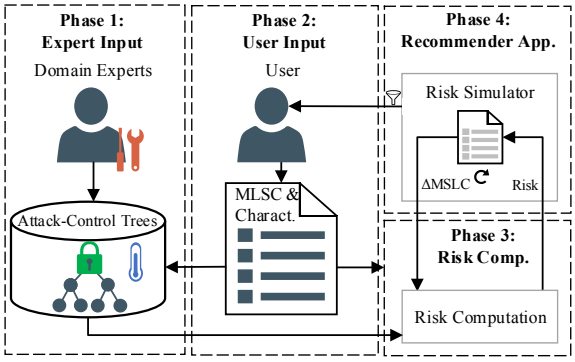


Figure 3.9: LiSRA: Overview[188]

inter-organizational security platform, and thus can benefit from the platform's update modules (cf. Fig. 3.6) reducing the need to bother the user with a long list of maturity levels to fill in.

Phase 3: Risk Computation The general risk computation process is illustrated in Fig. 3.10. Before the risk computation can start, the control dependencies are resolved. This is needed because many of the controls are dependent on each other so that their effect cannot be assessed independently [193]. Therefore, the effective maturity levels may be lower than the actual maturity levels due to control dependencies. Then, the total risk is derived from scenario risks that are calculated based on both the probability of adverse impact and its severity. The probability of adverse impact is the probability that an attack is initiated and succeeds. Both factors are calculated using attack trees and depend on

the chosen attacker type, e. g. a script kiddie goes for the cheapest attack while a nation-state attacker chooses the attack with the highest success-chances. The probability of attack success not only depends on the attacker, but also on the maturity level of assigned controls. Then, the probability of attack success is subsequently aggregated up the tree until the final attack goal is reached.

Phase 4: Recommender Application The next step, when the risk has been computed, is to identify the most beneficial security control to improve by increasing its maturity level. One option for that is to manually inspect the results of the risk analysis. If the total risk indicates the need for action one can go through the list of scenarios to find the high-risk scenarios and identify related security controls. However, LiSRA also offers a recommender application that automatizes the inspection process to find the most beneficial security control. By most beneficial, we mean the most effective and the most cost-efficient security control. Most existing approaches evaluate new security activities in isolation of security activities already in place, and they ignore that multiple overlapping activities can be of complementary, substitutive, or dependent nature which leads to an over-investment in security measures [21]. LiSRA explicitly addresses both aspects without bothering the user. Transparent recommendations are of crucial importance for the acceptance of recommender systems such as LiSRA. It describes to which extent users understand why a particular item is recommended to them [171]. Therefore, besides the recommendations themselves, also the rationale behind the recommendations is presented to the user by a graphical explanation interface.

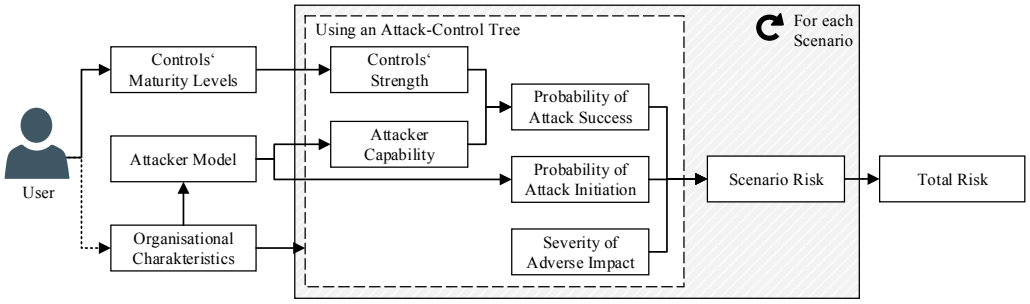


Figure 3.10: LiSRA: General Risk Computation Process [188]

LiSRA was implemented in Java and the evaluation showed it is robust against logical transformations of the underlying attack trees, with good performance, and perceived as useful by a focus group. However, it has some limitations. On the one hand, it does not consider adaptive or multiple-shot adversaries. On the other hand, it is particularly designed for small and medium enterprises, where a certain maturity level for a security control is consistent in the whole scope. Larger organizations might have different maturity levels for security controls in different security zones or even for different assets. Subsequent work adds to the visualization of the attack trees [190].

3.2 Security Risk Assessment for Large Enterprises

As we just have seen when discussing the limitations of LiSRA, small and medium enterprises might have different needs than large enterprises. Therefore, within this section, we consider two different challenges: In Sect. 3.2.1, a larger company with several subsidiaries wants to compare their subsidiaries' security levels. Section 3.2.2 discusses risk management of mobile devices, namely for smartphone apps.

3.2.1 Comparison of Subsidiaries' Security Levels in E-Commerce

In general, the comparison of subsidiaries' security levels could be done by treating them as different companies and applying a separate LiSRA instance for all of them. However, LiSRA focuses on risk assessment of a single enterprise along with recommendations for improvements and the basic problem of comparing subsidiaries' security levels is closer to a multicriteria decision making problem. A natural

approach for that is the analytic hierarchy process (AHP) [179, 180], which makes the prioritization of controls – implicitly already included in LiSRA – explicit. The AHP allows a structured comparison of security controls maturity levels and also provides a ranking. The AHP breaks down a complex evaluation problem into manageable sub-problems by using pairwise comparisons. The comparisons use a predefined scale to assign numerical values according to different preferences. Based on the pairwise comparisons, a square matrix is derived and its normalized eigenvector is used as numerical measure for the preferences.

	Priority	Company1	Company2	Company3	Company4	Company5
Comparison eCommerce	100.0%					
A.5 Information security policies	1.2%					
A.5.1.1 Policies for information security	75.0%	33.3%	8.3%	16.7%	33.3%	8.3%
A.5.1.2 Review of the policies for information security	25.0%	29.2%	7.8%	29.2%	29.2%	4.6%
A.6 Organization of information security	1.7%					
A.6.1.1 Information security roles and responsibilties	21.2%	22.2%	22.2%	11.1%	22.2%	22.2%
A.6.1.2 Segregation of duties	16.7%	23.5%	5.9%	23.5%	23.5%	23.5%
A.6.1.3 Contact with authorities	13.5%	22.2%	22.2%	22.2%	22.2%	11.1%
A.6.1.4 Contact with special interest groups	8.0%	19.4%	5.2%	19.4%	19.4%	36.7%
A.6.1.5 Information security in project management	14.5%	40.6%	5.7%	21.5%	10.7%	21.5%
A.6.2.1 Mobile device policy	11.1%	16.7%	16.7%	16.7%	16.7%	33.3%
A.6.2.2 Teleworking	15.0%	28.6%	7.1%	28.6%	28.6%	7.1%

Figure 3.11: AHP Applied to Security Controls in E-Commerce [187]

Again, we select security controls based on ISO/IEC 27001 along with the COBIT framework since this data is already available for all subsidiaries in the ISMS system of the large enterprise. To not further complicate the problem, we only consider e-commerce subsidiaries within the large enterprise to avoid the need of cross-domain comparisons [187]. Figure 3.11 shows a part of the evaluation. The priority column reflects the weight of each security control within each subgroup and for each subgroup as part of the whole. The different results in the company columns stems from different maturity levels for the security controls.

Since the maturity levels are determined per asset within the subsidiaries, for each security control there are multiple maturity levels. The most natural approach would be to extent the AHP by one level and consider assets as another subcategory of the security controls, which would also allow to prioritize them. However, the AHP only works with a fixed set of categories, leading to problems if only some companies have a certain asset, e. g. a file-server. One can solve the problem by only considering assets which are common in all subsidiaries. However, this would draw only a limited picture. As a solution, an asset class for all remaining, unspecified assets could be introduced, but then again, the question is which aggregation should be used?

Therefore, we investigated different aggregation types [186]. Unfortunately, the process of aggregating maturity levels is neither well documented nor comprehensively studied or understood (from a psychological perspective), so most of this labor is done by rule-of-thumb [205]. We investigated four aggregation types - namely the minimum, maximum, average and median - to compare their different potential impacts on decision making.

Regarding average and median, strengths and weaknesses have been discussed in scientific literature. Averages are strongly influenced by extreme values. Although, the scale has only a limited range from zero to five, in this context, this could lead to an over- or underestimation of control maturity. Minimum and maximum further alleviate potential misrepresentations of control maturity, as they provide the numerical range of scores and expose potential outliers [23].

It is also worth to briefly consider different optimization strategies information security managers might follow, if they knew the aggregation method: Using the minimum would reward improving only the worst values. Seen as weakest link of a chain, this could make sense in some scenarios. Using the maximum rewards improving only the best value or do nothing if it is already at five, which is probably not desirable. Using the average rewards improving any value, most likely the easiest or cheapest ones are increased first. Using the median may lead to a really two-fold security level with half of the services being insecure and half of the services being secure.

We also investigated the different aggregations with real world data as shown in Tab. 3.1. One can notice that the maximum differs most from all other aggregations. Considering also the different strategies, as a result, we recommend to use the average or minimum as aggregation method.

Aggregation/Proportion	Company1	Company2	Company3	Company4	Company5
Average	16.7% *(4.)*	15.4% *(5.)*	19.8% *(1.)*	18.3% *(3.)*	19.5% *(2.)*
Median	16.7% *(4.)*	16.3% *(5.)*	19.8% *(1.)*	18.8% *(2.)*	18.1% *(3.)*
Minimum	16.6% *(4.)*	14.6% *(5.)*	21.3% *(1.)*	18.7% *(2.)*	18.5% *(3.)*
Maximum	17.5% *(2.)*	15.6% *(5.)*	16.1% *(4.)*	16.2% *(3.)*	24.2% *(1.)*

Table 3.1: AHP Applied to Different Aggregation Types for Security Controls for Multiple Assets [186]

3.2.2 Security Risk Management for Smartphone Apps

With a raising number of apps for smartphones and a great diversity of developers ranging from spare time developers to large companies, it is more difficult than ever to assess the risk of a certain app. Despite approaches to raise their awareness [127], spare time developers, but also large enterprises rely on libraries from other parties, which often spy on the users. However, none of the app stores offers a dedicated security or privacy score for such apps. With security policies such as "bring your own device" (BYOD) the lines between personal use and use for work are blurred. BYOD is an attractive employee IT ownership model that enables employees to bring and use their personal devices in enterprises. It provides more flexibility and productivity for the employees, but may impose some serious privacy and security risks since personal matters are mixed with work. One of the arising problems of BYOD is that in order to benefit from it, the IT security governance may not be as strict as it could be for a smartphone only used for work. But even if users would accept allowed and blocked lists, the decision which apps to block would need to be made by the IT department. Decisions should be made as a trade off between the necessity of the app for business (or personal) purposes and the risk with regard to enterprise assets.

For that purpose, we propose Enterprise Smartphone Apps Risk Assessment (ESARA) as a novel framework aimed at supporting enterprises to protect their data against adversaries and unauthorized accesses [92]. ESARA makes use of different approaches from literature and combines them with the app behavior analyzer and the app perception analyzer [91] to get a more realistic and holistic picture of installed apps. Requirements for the development of ESARA were:

(E.1) Reuse of existing approaches

(E.2) Limiting the necessary effort (since there is a large number of apps)

(E.3) Scalability in the way that it should be easy to rely on external services and allowing several companies to share a same infrastructure

(E.4) Independence from app markets since even after several years none of them offers a decent security or privacy score

(E.5) Involving employees for feedback when using an app

(E.6) Involving employees for decisions

Figure 3.12 shows an overview of the proposed architecture for ESARA, which consists of three main modules: employee's smartphone, server and enterprise IT department. On the employee's device there is an app running that analyzes the behavior of a certain installed app and ultimately communicates the results to the employee (behavior analyzer). To respect the employees' privacy while trying to identify security intrusive apps only the apps' permission requests are analyzed which is not as intrusive as run-time monitoring, where one could conclude what an employee was doing. Furthermore, information is only sent to the server when confirmed by the user, optionally along with reviews regarding each privacy and security invasive activity that the employees observe. The behavior analyzer also stores the employee's security and

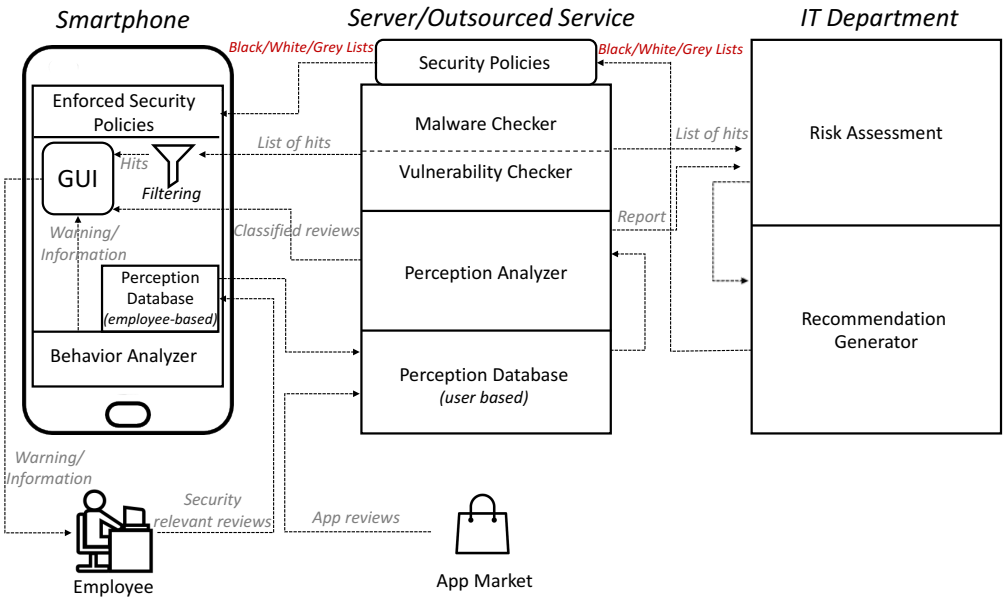


Figure 3.12: ESARA: Architecture Overview [92]

privacy perception in the perception database and receives results regarding the perception analysis and risk assessment from the other two modules.

The server or an outsourced service is supposed to check apps for vulnerabilities and malicious activities by running a malware and vulnerability scanner, therefore, it does not collect any data from employees and also avoids to deploy these checkers on resource constrained smartphones [32]. Diverse vulnerability checkers are available and fulfill the requirements [123, 124, 141, 142, 173]. This server/service is also responsible to analyze employees' and other users' perception (e. g. from the app stores) about security and privacy behavior of apps (perception analyzer). For that purpose, natural language processing (NLP) techniques (e. g. tokenization, stemming and removing stop words) along with sentiment analysis techniques are used to find both positive and negative reviews with regards to privacy and security aspects. If security policies are put in place, black, white and gray lists can also be distributed via this server/service.

The enterprise IT department takes the final decisions about which app is to place on which list – either manually or automatically by defining certain rule sets. For that purpose, reports written by the employees along with the automated analyses from the perception analyzer are used.

Table 3.2: Coverage of Top 10 Mobile App Risks [169] by ESARA

No.	Risk	Malware Checker	Vuln. Checker	Behavior Analyzer	Perception Analyzer
1	Activity monitoring and data retrieval	✓	–	(✓)	(✓)
2	Unauthorized dialing, SMS, and payments	✓	–	✓	(✓)
3	Unauthorized network connectivity	(✓)	–	–	(✓)
4	UI Impersonation	–	–	–	(✓)
5	System modification	✓	–	–	✓
6	Logic or Time bomb	✓	(✓)	–	–
7	Sensitive data leakage	(✓)	✓	✓	✓
8	Unsafe sensitive data storage	–	✓	–	(✓)
9	Unsafe sensitive data transmission	–	✓	–	✓
10	Hardcoded password/keys	✓	✓	–	–

For the evaluation of ESARA, besides the evaluation of the behavior and perception analyzer, we checked its coverage of the most prevalent mobile app risks taken from Veracode's [169] top 10 mobile app risks and investigated the robustness of ESARA in assessment and detection of each individual risks. Table 3.2 also connects ESARA's components with the mobile app risks, demonstrating that ESARA covers all the listed risks and nearly all of them by at least two components. Besides that, all requirements we defined afore are considered. As future work, a real implementation along with testing outside of a laboratory environment and user studies to investigate the users' acceptance are planned.

3.3 Cloud Service Provider Security for Customers

Cloud Computing has been emerging as the new computing paradigm in the last ten years, enabling consumers to flexibly purchase computing power and storage capacity on-demand, conveniently and cost efficiently from specialized providers. Recent studies claim that cloud computing has left the hype phase behind and can already be considered the norm for IT [27]. However, besides the potential economic benefits of cloud adoption, there are also security concerns as it represents a form of IT outsourcing and exhibits technological peculiarities concerning size, structure and geographical dispersion [120]. As a consequence, cloud customers are often afraid of loosing control over their data and applications and of being exposed to data loss, data compliance and privacy risks. On the other hand, there may be also benefits to security in the cloud, since a cloud service provider (CSP) enjoys economies of scale in terms of security as well, being able to invest more and thereby achieve a higher security level on a much larger scale than most client companies would with an in-house data center [77, 109].

So one would expect, that a cloud customer will most likely engage with a CSP demonstrating a high level of security. However, in practice there are two challenges. First, selecting the most secure CSP is not straightforward. With the outsourcing the customer also delegates the implementation of security controls to the CSP. From a CSP's view, its main objective is to make profit. Therefore, it can be assumed that the CSP does not want to invest more than necessary in security. Additionally to the different objectives of customer and CSP, there is the problem that security – compared to other providers' attributes like cost or performance – is not easily measurable and there are no precise metrics to quantify it [25].

The consequences are twofold. It is not only hard for the tenant to assess the security of outsourced services, it is also hard for the CSPs to demonstrate their security capabilities. Even if a CSP puts a lot of effort in security, it will be hard to demonstrate it to the customer, since malicious CSPs will pretend to do the same. This imbalance of knowledge is long known as information asymmetry [6] and together with the cost of cognition to identify a good provider and negotiate a contract [207] has been widely studied in economics.

The contribution of this section is twofold. First, we will present an investigation how companies choose as CSP [155] and propose a method to support the selection of a secure provider [161]. Second, we propose a model to support the systematic analysis of attacks on cloud customers [24].

3.3.1 Secure Cloud Provider Selection

In this section, we first report about the investigation of the role of security in cloud service provider selection. We then briefly describe the Consensus Assessment Initiative Questionnaire (CAIQ), a questionnaire from the Cloud Security Alliance (CSA) to determine the security of CSPs. Based on the questionnaire we then propose a method to compare the security of multiple CSPs.

Decisive Factors in Cloud Service Provider Selection

We investigated organizations' practices when selecting a CSP and expected to verify the importance of security. Furthermore, we expected customers as well as CSPs to come up with security assurance methods to verify and respectively demonstrate their security efforts. For that purpose, we interviewed practitioners from eight German companies who deal with CSP selection [155].

The respondents were asked about criteria and requirements for the selection of a CSP instead of directly asking them about the role of security. While security was rarely mentioned first, it was sooner or later addressed in all the discussions. Mentioned selection criteria were costs, size of provider, and by ease of use.

Others already pointing in the direction of security were trust, compliance, and confidentiality of their users' data. The participants also gave insights into processes of CSP selection within the organizations. Some were using multiple CSPs and choosing them per project or task, for some the provider decision was made on a higher hierarchical level, and several respondents admitted that the choice for a CSP was made by chance, e. g. simply choose any convenient provider to make the first steps in the cloud, because a developer already had some experience with it or just because the company had a voucher.

We further investigated the moderate interest in security and found that most respondents were more focused on mitigating risks, e. g. by regarding the location of a provider as an indication for trustworthiness or considering the criticality of data placed in the cloud in relation to the security level. Further answers revealed that some respondents were assuming that many users trust their providers without any proof, in particular when sticking with a large CSP such as Amazon, they referred to the "IBM Effect" stating that "No one ever got fired for buying IBM" applies to Amazon's AWS nowadays. This supports the assumption that the requirement on security is extrinsically motivated by compliance.

Another objective was to gain some insights whether and how the respondents verified the security levels of their CSPs. Here the respondents mostly named non-technical measures such as certification, (financial) audits checking for the capability of a CSP to grant compensations, and contractual agreements. Besides that, few respondents named security tests and two also presented their own risk evaluation respectively questionnaire for the CSP. However, several respondents also expressed skepticism when talking about assurance, criticizing external auditors or service level agreements as toothless and pointing out that the need to control or verify everything, although one had outsourced, is unnecessary costly.

In summary, the collected findings on the role of security in CSP selection were ambiguous. Security however, was never the first answer of the respondents and most of them could not provide specific security requirements. On the other hand, security as a requirement was present in all the discussions, and showed up particularly as availability and in rare cases as confidentiality. In the investigated sample we could rarely find any elaborated process of eliciting requirements and then coming to a rational decision which CSP to select. Instead, CSPs were chosen based on vouchers, by chance, or by the management because of established relationships. Another identified pattern was that companies often try to 'first get into the cloud' and then optimize costs and sometimes security (lift and shift). In all phases of the selection, the requirement elicitation, the decision making process and in the use of assurance technologies there seems to be a gap between research and practice. This gap is quite common in a lot of areas [139].

Consensus Assessments Initiative Questionnaire

It seems that questionnaires to the CSPs are the only way of gathering information on the security of a CSP, in particular before there is a business relationship established, which might allow to test certain security parameters. An obvious strategy for the cloud customer is to ask the CSP to answer a set of questions from a proprietary questionnaire and then try to fix the most relevant issues in the service level agreements. However, this makes the evaluation process inefficient and costly for the customers and the CSPs.

To standardize the requests and render them unnecessary the Cloud Security Alliance [41], a non-profit organization with the aim to promote best practices for providing security assurance within cloud computing, has provided the Cloud Controls Matrix (CCM) and the Consensus Assessments Initiative Questionnaire (CAIQ). The CCM [39] is designed to guide cloud vendors in improving and documenting the security of their services and to assist potential customers in assessing the security risks of a CSP.

Each control consists of a control specification which describes a best practice to improve the security of the offered service. These controls are mapped to other industry-accepted security standards, regulations, and controls frameworks, e. g. ISO/IEC 27001/27002/27017/27018, NIST SP 800-53, PCI DSS, and ISACA COBIT.

For each control in the CCM the CAIQ [38] contains one or more associated 'yes or no' questions asking if the CSP has implemented the respective control (see Tab. 3.3 for an overview of the CAIQ's structure and Fig. 3.13 for some example questions).

Control Domain	Control ID	Question ID	Control Specification	Consensus Assessment Questions	Consensus Assessment Answers			Notes
					Yes	No	Not Applicable	
Application & Interface Security *Application Security*	AIS-01	AIS-01.1	Applications and programming interfaces (APIs) shall be designed, developed, deployed, and tested in accordance with leading industry standards (e.g., OWASP for web applications) and adhere to applicable legal, statutory, or regulatory compliance obligations.	Do you use industry standards (i.e. OWASP Software Assurance Maturity Model, ISO 27034) to build in security for your Systems/Software Development Lifecycle (SDLC)?				
		AIS-01.2		Do you use an automated source code analysis tool to detect security defects in code prior to production?				
		AIS-01.3		Do you use manual source-code analysis to detect security defects in code prior to production?				
		AIS-01.4		Do you verify that all of your software suppliers adhere to industry standards for Systems/Software Development Lifecycle (SDLC) security?				
		AIS-01.5		(SaaS only) Do you review your applications for security vulnerabilities and address any issues prior to deployment to production?				
Application & Interface Security	AIS-02	AIS-02.1	Prior to granting customers access to data, assets, and information systems, identified security, contractual, and regulatory requirements for customer access shall be	Are all identified security, contractual, and regulatory requirements for customer access contractually addressed and remediated prior to granting customers access to data, assets, and information systems?				
		AIS- 02.2		Are all requirements and trust levels for customers' access defined and documented?				
Application & Interface Security *Data Integrity*	AIS-03	AIS-03.1	Data input and output integrity routines (i.e., reconciliation and edit checks) shall be implemented for application interfaces and databases to prevent manual or systematic processing errors, corruption of data, or misuse.	Does your data management policies and procedures require audits to verify data input and output integrity routines?				
		AIS-03.2		Are data input and output integrity routines (i.e. MD5/SHA checksums) implemented for application interfaces and databases to prevent manual or systematic processing errors or corruption of data?				
Application & Interface Security *Data Security / Integrity*	AIS-04	AIS-04.1	Policies and procedures shall be established and maintained in support of data security to include (confidentiality, integrity, and availability) across multiple system interfaces, jurisdictions, and business functions to prevent improper disclosure, alteration, or destruction.	Is your Data Security Architecture designed using an industry standard (e.g., CDSA, MULITSAFE, CSA Trusted Cloud Architectural Standard, FedRAMP, CAESARS)?				
Audit Assurance & Compliance *Audit Planning*	AAC-01	AAC-01.1	Audit plans shall be developed and maintained to address business process disruptions. Auditing plans shall focus on reviewing the effectiveness of the implementation of security operations. All audit activities must be agreed upon prior to executing any audits.	Do you develop and maintain an agreed upon audit plan (e.g., scope, objective, frequency, resources,etc.) for reviewing the efficiency and effectiveness of implemented security controls?				
		AAC-01.2		Does your audit program take into account effectiveness of implementation of security operations?				
Audit Assurance & Compliance *Independent Audits*	AAC-02	AAC-02.1	Independent reviews and assessments shall be performed at least annually to ensure that the organization addresses nonconformities of established policies, standards, procedures, and compliance obligations.	Do you allow tenants to view your SOC2/ISO 27001 or similar third-party audit or certification reports?				
		AAC-02.2		Do you conduct network penetration tests of your cloud service infrastructure at least annually?				
		AAC-02.3		Do you conduct application penetration tests of your cloud infrastructure regularly as prescribed by industry best practices and guidance?				
		AAC-02.4		Do you conduct internal audits at least annually?				
		AAC-02.5		Do you conduct independent audits at least annually?				
		AAC-02.6		Are the results of the penetration tests available to tenants at their request?				
		AAC-02.7		Are the results of internal and external audits available to tenants at their request?				

Figure 3.13: Consensus Assessments Initiative Questionnaire in Version 3.1 [39]

Table 3.3: CCM-Item and CAIQ-Question Numbers per Domain (version 3.1) [161]

ID	Domain	CCM	CAIQ
AIS	Application & Interface Security	4	9
AAC	Audit Assurance & Compliance	3	13
BCR	Business Continuity Management & Operational Resilience	11	22
CCC	Change Control & Configuration Management	5	10
DSI	Change Control & Configuration Management	7	17
DCS	Datacenter Security	9	11
EKM	Encryption & Key Management	4	14
GRM	Governance and Risk Management	11	22
HRS	Human Resources	11	24
IAM	Identity & Access Management	13	40
IVS	Infrastructure & Virtualization Security	13	33
IPY	Interoperability & Portability	5	8
MOS	Mobile Security	20	29
SEF	Security Incident Management, E-Discovery & Cloud Forensics	5	13
STA	Supply Chain Management, Transparency and Accountability	9	20
TVM	Threat and Vulnerability Management	3	10
	Total	133	295

Selection of a Secure Cloud Service Provider

Overall, the CAIQ (in version 3.1) contains 295 questions. As an experiment, we asked participants to decide for an imaginary scenario which out of two CSPs offers the more suitable security [161]. In order to keep the experiment manageable, we only gave the participants a small subset (20 questions and answers) of the CAIQ. While most of them were able to correctly identify the more suitable CSP, the participants were not confident about the ease of use and usefulness of the manual approach. Mainly, because even if they worked only on a subset, the imagination of doing the comparison with the full set of 295 questions, identifying the related questions and comparing the results seemed cumbersome to them.

Given that in practice more than 2 CSPs need to be compared, a more automatic approach is necessary. As of March 2020, the Cloud Security Alliance listed 733 CSPs with 690 CAIQs and 106 certifications[2].

[2]Note that some companies list the self-assessment along with their certification, some do not provide their self-assessment when they got a certification.

For that purpose, we developed an approach that facilitates the comparison of the security posture of CSPs based on answers to the CAIQ (cf. Fig. 3.14). The three main actors involved are the tenant, the alternative

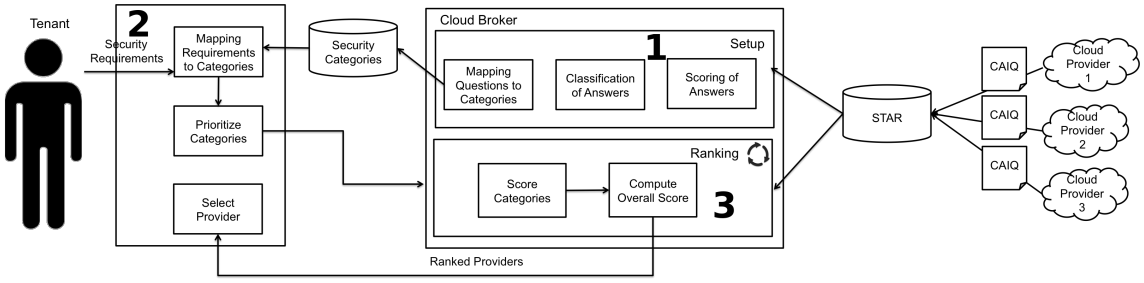


Figure 3.14: CPS [161]

CSPs, and a cloud broker. A cloud broker is an intermediary between the CSPs and the tenant helping the tenant to choose a provider tailored to his (security) needs (cf. NIST Cloud Computing Security Reference Architecture [74]). The suggested approach consists of three phases:

1. In the setup, the broker has to assess the answers of the CSPs to the CAIQ (classification and scoring) and defines security categories which are mapped to the CAIQ's questions. The list of security categories is then provided to the tenant.

2. The tenants map their security requirement to the security categories provided by the broker and prioritize them. The tenants also need to provide a (rough) description of their service requests or lists of CSPs they want to compare.

3. The broker first selects candidate CSPs delivering the services requested by the customers or uses the provided list for a start. The broker ranks the candidate providers then based on the prioritization of security categories specified by the customers and the answers that the CSPs gave to the CAIQ. The list of ranked CSPs is then returned to the customers, who can use the list as part of their selection process, i. e. by manually comparing the top 5 candidates or using the result of the security comparison as a building block where other factors such as costs and performance are also considered.

The approach to rank CSPs adopts the Analytic Hierarchy Process (AHP) [179] similar to what we have described in Sect. 3.2.1 already. In the same manner, the output of the presented approach is a hierarchy where each CSP gets a overall score and a score for each security category, allowing the customer not only to use the overall result of the ranking, but also to reproduce each CSP's strengths and weaknesses. This allows the customers further reasoning or an adaptation of the requirements/scoring should they not be confident with the result.

The presented approach is the first approach for CSP selection with an effective way to measure and compare the security of a provider. Previous works have considered security as a relevant criteria for the comparison and ranking of CSPs [43, 66, 70, 76, 164, 203, 216]. However, most of the approaches did not suggest a method for the collection of data about the CSPs' security. Closest to the presented approach is the approach by Habib et al. [76], which identified CAIQ as data source, but did not specify in which way the data should be used. The proposed approach could be used as a building block for the existing approaches to CSP selection that consider also other providers' attributes like cost and performance.

3.3.2 Supporting Security Assessments

In this section, we introduce a high level approach to support cloud customers in their security assessments of the clouds [24]. The idea is to capture the security requirements of cloud customers as well as characteristics of attackers. The model can be used for deriving new security threats from existing scenarios, as well as describing and analyzing new what-if scenarios by changing characteristics of involved parties.

System Model

We define a model of a cloud environment on an Infrastructure-as-a-Service layer consisting of entities and the system components as shown on Figure 3.15.

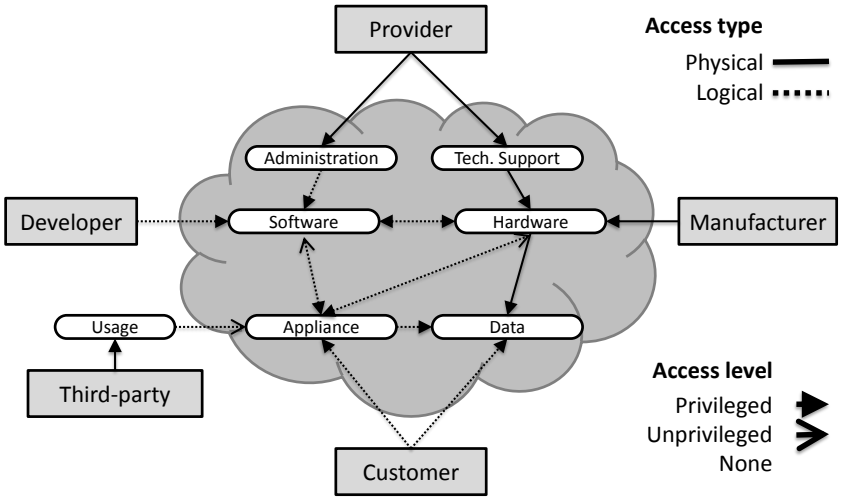


Figure 3.15: System Model with Relations Between Entities and Components [24]

Entities represent subjects which are involved in a cloud service, directly or indirectly, while components represent objects of which a cloud service is composed of. The entities include: a cloud service provider who manages and operates a cloud infrastructure, which includes hardware and software resources; the manufacturer who produces the hardware resource used by the provider; a developer who produces the software resource used by the provider; the customer who uses the cloud service; and third parties which are not directly involved in providing or using Infrastructure-as-a-Service, but can represent user on higher layers of the cloud service (e. g. Software-as-a-Service). Each entity has one or more components, which can be accessed physically or logically, e. g. the provider has an administration maintaining the software (logical access) and a technical support team maintaining the hardware (physical access). Each entity or component can have multiple instances when used for describing an attack scenario, e. g. there can be several customers.

The relationship between entities and their components, as well as between components themselves, is defined through different access levels: privileged means full access with all the privileges for configuring and manipulating a component; unprivileged means limited access to functionality or an interface of a component; and none means no access at all. Access levels are directed and transitive: A can use its access to B in order to manipulate C, when B has access to C.

Different archetypes describe the contributors to an attack: malicious (intentionally contributing to an attack); ostrich (knowingly contribute to an attack); charlatan (failing to acquire essential knowledge about contributing to an attack); stepping stone (unknowingly contributing to an attack). The malicious and ostrich archetypes are driven by goals, e. g. causing damages or for monetary reasons, and their skill level determines the success of reaching such goals. The charlatan and stepping stone archetypes have low skills, which renders their goal of providing a secure cloud service to their customers unsuccessfully. The ostrich can also been called lazy, and the term sloppy can been used for charlatans and stepping stones.

Evaluation of the System Model

We evaluated the system model by applying it to already known attacks and investigating its modeling ability. In this case, we consider a side-channel attack. The setup of a side-channel attack scenario consists of a customer who tries to attack another customer by placing a virtual machine on the same physical server and by trying to observe the system's behavior [178]. In this case almost all entities are involved as shown in Fig. 3.16.

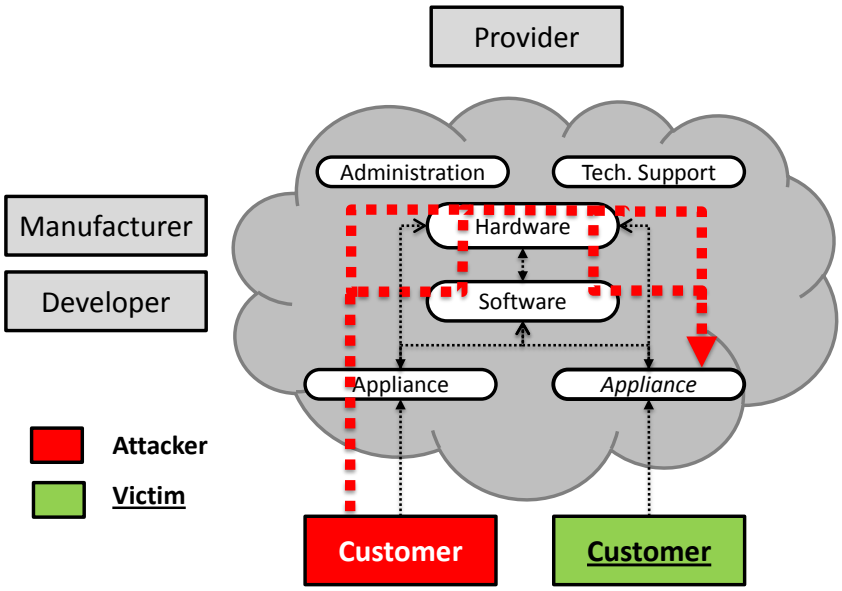


Figure 3.16: Attacking Other Customers Through Side-channels in Hardware and/or Software [24]

The CSP configures and chooses the hardware and software (operating system, hypervisor, etc.) which are supplied by the manufacturer and the developer, respectively. The input of the manufacturer and the developer depends on their archetypes. In this scenario it is not reasonable to consider them being malicious, but the remaining range from ostrich to defender may result in input from low quality hardware / software to specially hardened ones counteracting side-channel attacks. The CSP also influences the feasibility of side-channel attacks, since he configures the system and has to justify his choices of the used software and hardware.

In the considered side-channel attacks, one customer (red) attacks another customer (green) by using his appliance to observe characteristics of the hardware directly or via the software. The attacker tries to gather information by eavesdropping on the data processed in the attacked appliance of the other involved customer. The attacked customer can hardly do anything to protect himself against side-channel attacks besides paying to use physical resources exclusively. However, if the CSP is a defender, the CSP can monitor appliance integrity from the software in order to protect the customer [9, 65], provide recovery options once intrusion has been detected and removed [114] or install a secured environment like SICE [10].

Since we positively evaluated the system model with three more attacks, and we were also able to construct four more “what-if”-scenarios constructing theoretical, new attacks, we conclude that the proposed system model is helpful for a customer in assessing cloud computing security. In particular, the system model helps identifying characteristics for the involved service providers which can aid to attacks of other parties.

Chapter 4

Privacy Enhancing Technologies

> If you care about privacy online, you need to actively protect it.
>
> *Roger Dingledine*

Bruce Schneier states [131]: ”Surveillance is the business model of the internet. Everyone is under constant surveillance by many companies, ranging from social networks like Facebook to cellphone providers.“ One of the reasons for surveying user is a rising economic interest in the internet [20]. However, users are not helpless and can make use of privacy-enhancing technologies (PETs) to protect them. Examples of PETs include services which allow anonymous communication, such as Tor [206] or JonDonym [110].

Tor and JonDonym are low latency anonymity services which redirect packets in a certain way to hide metadata (the sender's and optionally – in case of a hidden service – the receiver's internet protocol (ip) address) from passive network observers. While Tor and JonDonym differ technically, they are highly comparable with respect to the general technical structure and the use cases. Tor offers an adapted browser including the Tor client for using the Tor network, the “Tor Browser”. Similarly, the “JonDoBrowser“ includes the JonDo client for using the JonDonym network.

However, the entities who operate the PETs are different. Tor is operated by a non-profit organization with thousands of voluntarily operated servers (relays) and an estimated 2 million daily users by the Tor Project [206] and an estimated 8 million daily users by Mani et al. [126]. Tor is free to use with the option that users can donate to the Tor project. JonDonym is run by a commercial company with servers (mix cascades) operated by independent and non interrelated organizations or private individuals who all publish their identity. A limited service is available for free, and different premium rates allow to overcome the limitations. The actual number of users is not predictable since the service does not keep track of this.

However, while the user number of anonymization services is large enough to conduct studies and evaluate the running systems, it is quite low compared to the number of internet users in total, which was estimated to 4.13 billion in 2019 [35]. Far less than 1% of the users use anonymization networks.

In order to investigate why there isn't a broader adoption of anonymization services, some retrospective requirements engineering seems to be necessary: Investigating users privacy concerns and their technology acceptance to find factors promoting the use of PETs. Since Tor is one of the most prominent PETs, the hope is that the insights can also be transferred to other PETs.

Besides the users' perspective, it is also important to investigate the economic side: Are users willing to pay for PETs and which incentives and hindrances exist for companies to implement PETs?

Besides stand-alone PETs, another way to promote privacy is to integrate it in existing services or design services with privacy in mind (privacy by design). The last part therefore deals with the application of privacy by design for online shopping and the internet of things.

The remainder of this chapter is structured as follows:

- Sect. 4.1 discusses how the distribution of PETs could be increased but investigating user's concerns, technology acceptance and willingness to pay as well as business models on PETs.

  - Sect. 4.1.1 investigates technology use factors for the anonymization networks Jondonym [79, 80] (cf. Sect. C.1, C.4) and Tor [83, 84] (cf. Sect. C.8, C.9) and compares them [90] (cf. Sect. C.10).
  - Sect. 4.1.2 assesses incentives for customers to pay for PETs [89] (cf. Sect. C.7) as well as incentives and barriers for companies to build a business model on PETs [88] (cf. Sect. C.2).
- Sect. 4.2 discusses application of privacy by design.
  - Sect. 4.2.1 discusses different architectures for pseudonymous online shopping [157] (cf. Sect. C.3).
  - Sect. 4.2.2 investigates privacy by design in the internet of things by investigating privacy policies [165] (cf. Sect. C.5) and privacy patterns [154] (cf. Sect. C.6).

The respective papers can be found in Appendix C and the author's contribution for each paper is indicated in Tab. **??** on page **??**.

4.1 Users' Technology Acceptance and Economic Incentives

For PETs like anonymization networks like Tor [206] or JonDonym [110] which allow anonymous communication, there has been a lot of research [133, 181], but the large majority of it is of technical nature and does not consider the users and their perceptions. However, the number of users is essential for anonymization networks since an increasing number of (active) users also increases the anonymity set. The anonymity set is the set of all possible subjects who might be related to an action [168], thus a larger anonymity set may make it more difficult for an attacker to identify the sender or receiver of a message. Therefore, it is crucial to understand the reasons for the users' intention to use a PET or obstacles preventing it [3].

However, for the distribution of a PET it is not only important to understand the users' intentions to use the PET, but also the users' willingness to pay for the service, which would allow companies to build a business model upon the provision of the service. The main challenge in motivating the user to pay for an anonymization service is that the user can barely notice a working PET like an anonymization network directly. Noticing it is in most cases the result of a limitation of throughput, performance, or response time. Indirect effects such as fewer profiling are also hard to detect, but even harder to connect to a PET in place. This makes it hard for a company as well as the user to sell or, respectively, understand the advantages for these types of PETs. As a consequence, it is hard for a company to come up with a business model, and thus the further distribution of PETs is prevented.

Therefore, besides investigating the users' intention to use a PET in Sect. 4.1.1, we also investigate in Sect. 4.1.2 the economic sides of PETs from the perspective of the users' willingness to pay and from the perspective of a business owner to provide a PET as service.

4.1.1 User Concerns and Technology Acceptance Models

To investigate the users intention to use Tor or JonDonym we made us of two different popular structural equation [78] models:

Internet Users' Information Privacy Concerns (IUIPC) is a construct by Malhotra et al. [125] for measuring and explaining privacy concerns of online users. IUIPC is operationalized as a second-order construct[1] of the sub-constructs collection, awareness and control. That means the user's concerns are determined by concerns about data on the user in relation to the value or received benefits, by concerns about the control users have over their own data, and by concerns about his or her awareness regarding organizational privacy practices. IUIPC then influences trusting beliefs and risk beliefs which then influence the user's behavior, which was in the original research the release of personal information to a marketing service provider. The trusting and risk beliefs refer to the users' perceptions about the behavior of online firms (in general) to protect or lose the users' personal information.

[1] For an extensive discussion on second-order constructs see Steward [202].

Technology Acceptance Model (TAM) was developed by Davis [44, 45] based on the the theory of reasoned action (TRA) by Fishbein and Ajzen [63] and the theory of planned behavior (TPB) Ajzen [5]. According to the TRA, a person's behavioral intention determines that persons behavior. The behavioral intention itself is influenced by the person's subjective norms and attitude toward the behavior. The subjective norms refer to a person's normative beliefs and normative pressure to perform or not perform the behavior. The attitude relies on the person's beliefs about the behavior and its consequences. TPB is an extension of the TRA with the same overall structural process: the behavioral intention is influenced by several components and influences the behavior. However, the TPB adds perceived behavioral control which refers to a person's perception regarding the ease or difficulty of performing a given behaviour in a given situation.

Internet Users Information Privacy Concerns

We conducted a survey among users of the anonymization services JonDonym (141 valid questionnaires [80, 85]) and Tor (124 valid questionnaires [83, 86]) to investigate how the users' privacy concerns influence their behavioral intention to use the service.

For that purpose we used the IUIPC construct [125, 159, 160]. The IUIPC construct has been used in various contexts, such as internet of things [136], internet transactions [95] and mobile apps [174], but so far it had not been applied to a PET such as an anonymization service. There is a major difference between PETs and the other services regarding the application of the IUIPC instrument. The other services had a certain use for their customer (primary use) and the users' privacy concerns were investigated for the use of the service. The concepts of trusting and risk beliefs matched that in a way that they were referring to 'general companies' which may provide a service to the user based on data they receive. However, for anonymization services providing privacy is the primary purpose. Therefore, it is necessary to distinguish between trusting and risk beliefs with respect to technologies which aim to protect personal data (PETs) and regular internet services. As a consequence, the trust model within IUIPC's causal model was extended by trusting beliefs in Tor/JonDonym.

We tested the model using SmartPLS version 3.2.6 [177]. The measurement model was consistent and checks were fine for reliability and validity on both data sets. Figure 4.1 shows the structural equation model for Jondonym users and Fig. 4.2 for Tor users.

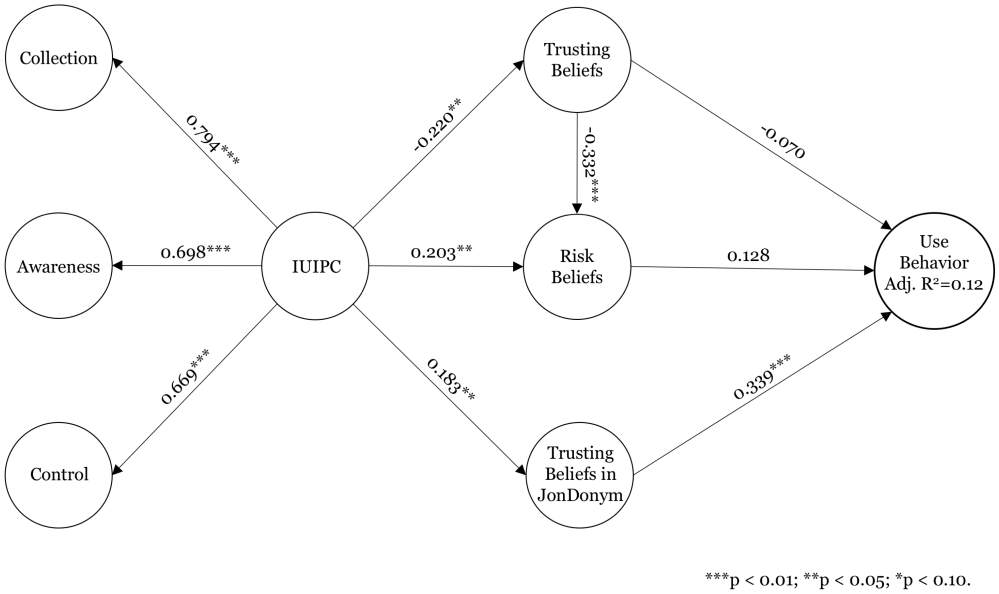


Figure 4.1: JonDonym Users, IUIPC, Path Estimates and Adjusted R^2-values of the Structural Model [80]

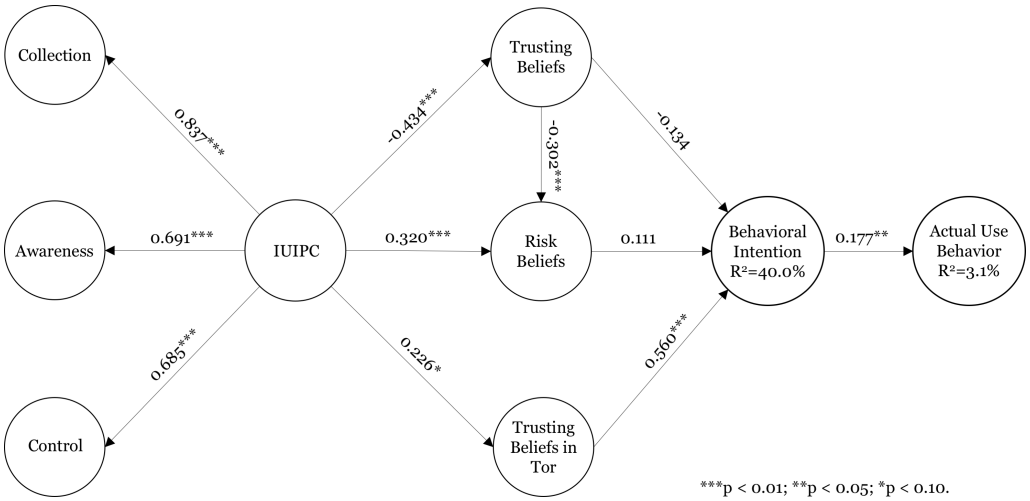


Figure 4.2: Tor Users, IUIPC, Path Estimates and Adjusted R^2-values of the Structural Model [83]

The models for JonDonym and Tor users turned out to be very similar. Most of the relations were as expected, somewhat surprising was the result that general trusting and risk belief had no significant effect on the use behavior. However, for the rather small effect sizes, it might be that the sample size was simply not large enough to show a significant relationship. In any case the trust in JonDonym or Tor had by far a larger influence on the use behavior, respectively the behavioral intention. The result shows that the reputation of being a trustworthy provider respectively service is crucial for an anonymization service provider. The results also show that users with a higher level of privacy concern rather tend to trust their anonymization service provider, which might be affected by the fact that we only asked users of the respective PET.

In general, if there is a reliable measure of the use behavior, this is a better indicator than the users' behavioral intention to use a service. Since we questioned actual users, we could use their use frequency of the services. However, it showed for Tor that the influence of the behavioral intention on the actual use behavior was rather small.

Users' attitudes and behavioral intention often differ from the behavior decisions they make often denoted as 'privacy paradox' [67]. Two possible explanations come to mind to explain the privacy paradox: i) users balance between potential risks and benefits they gain from the service (privacy calculus) [54], ii) users are concerned but lack knowledge to react in a way that would reflect their needs [208]. However, since we surveyed active users of Tor, both argumentations do not fit. Regarding the privacy paradox, we have already discussed how PETs differ from regular internet services. Regarding the lack of knowledge, users have already installed the PET and use it. However, it is still important to investigate the users' capabilities since users need a certain amount of knowledge in order to adequately evaluate the given level of privacy [163, 208]. For that purpose, we added the users' privacy literacy measured with the "Online Privacy Literacy Scale" (OPLIS) [128] to the model. It showed that users' privacy literacy positively influence trusting beliefs in Tor (cf. Fig. 4.3). Therefore, educating users and increasing their privacy literacy should add to the behavioral intention of using Tor. We will further investigate the influence of the behavioral intention on the actual use behavior by making use of the TAM model in the next subsection.

Technology Acceptance Models

Within the same survey, we also asked the participants about certain constructs we could use in a TAM model [82]: How they perceived the usefulness, the ease of use and the anonymity of the PET. Since we had already identified trust in the PET as a major driver for the behavioral intention, we included it too. The resulting model is shown in Fig. 4.4 including JonDonym and Tor users [90]. The model shows significant relationships for all paths as already known from the TAM model with three noteworthy observations:

- There are three main drivers of the PETs' perceived usefulness: perceived anonymity, trust and perceived ease of use which explain almost two-thirds of its variance. This demonstrates that for PETs

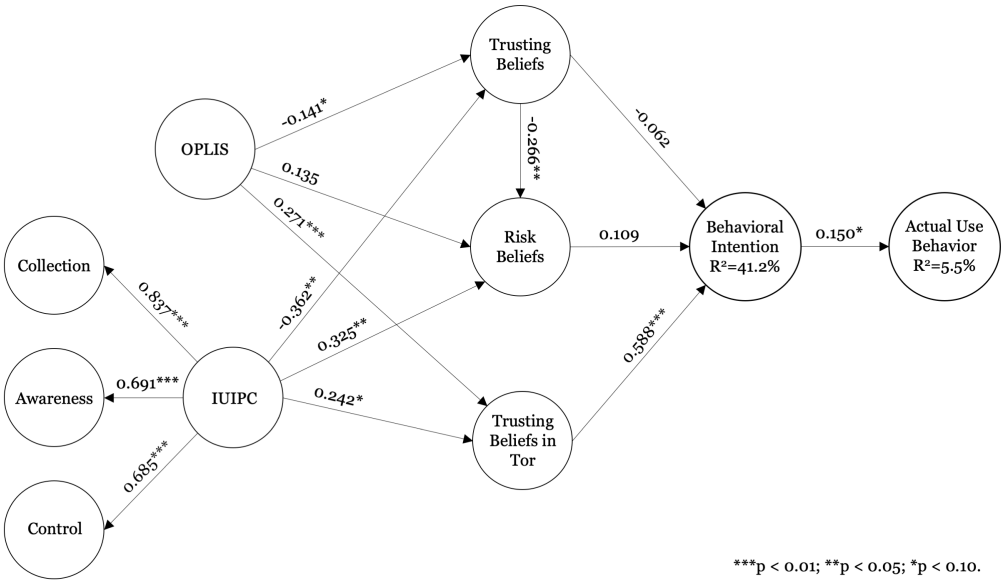


Figure 4.3: Tor Users, IUIPC & OPLIS, Path Estimates and Adjusted R^2-values of the Structural Model [84]

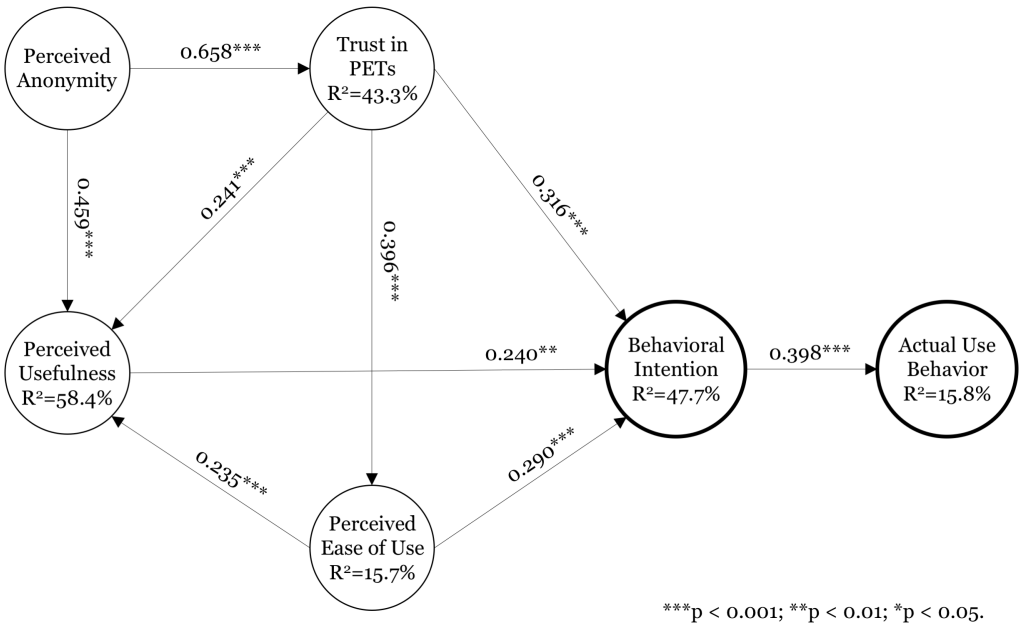


Figure 4.4: Tor/Jondonym Users, TAM, Path Estimates and Adjusted R^2-values of the Structural Model [90]

the two newly added variables perceived anonymity and trust in the PETs can be important antecedents in an technology acceptance models for PETs.

- Similar than in the IUIPC model, trust in the PET is the most important factor for behavioral intention. One more time emphasizing trust in the PETs as a highly relevant concept when determining the drivers of users use behavior of PETs.

- Since the effects of perceived anonymity and trust in the PETs on behavioral intention and actual use behavior were partially indirect, we calculated the total effects. Table 4.1 shows that the total effects for behavioral intention are relatively large and highly statistically significant.

Table 4.1: Tor and Jondonym Users, TAM, Total effects [90]

Total effect	Effect size	P-value
PA → BI	0.446	< 0.001
PA → USE	0.177	< 0.001
Trust_{PETs} → BI	0.511	< 0.001
Trust_{PETs} → USE	0.203	< 0.001

BI: Behavioral Intention PA: Perceived Anonymity USE: Actual Use Frequency

To investigate the differences between JonDonym and Tor and also to further investigate the small effect of behavioral intention on actual use behavior, we conducted a multigroup analysis to test whether there are statistically significant differences between JonDonym and Tor users as shown in Tab. 4.2.

Table 4.2: Tor and Jondonym Users, TAM, Multi-Group Analysis [90]

Relationships	Path coeff. original (JonDonym)	Path coeff. original (Tor)	P-values (JonDonym)	P-values (Tor)	Difference path coeff. (JonDonym vs Tor)	P-values
PA → Trust_{PETs}	0.597	0.709	< 0.001	< 0.001	0.112	0.865
PA → PU	0.543	0.369	< 0.001	< 0.001	0.174	0.088
Trust_{PETs} → BI	0.416	0.232	< 0.001	0.010	0.184	0.064
Trust_{PETs} → PU	0.173	0.304	0.035	0.008	0.131	0.823
Trust_{PETs} → PEOU	0.378	0.431	< 0.001	< 0.001	0.053	0.657
PU → BI	0.183	0.300	0.046	0.002	0.117	0.805
PEOU → BI	0.206	0.371	0.011	< 0.001	0.165	0.929
PEOU → PU	0.182	0.300	0.039	< 0.001	0.118	0.830
BI → USE	0.679	0.179	< 0.001	0.029	0.500	**< 0.001**

BI: Behavioral Intention PEOU: Perceived Ease of Use PA: Perceived Anonymity USE: Actual Use Frequency
PU: Perceived Usefulness of Protecting Users' Privacy

It showed that the most significant difference between JonDonym and Tor users was the effect size between behavioral intention and actual use, which is 0.679 for JonDonym and 0.179 for Tor. Less significant observations were that the effects of trust on behavioral intention and perceived anonymity on perceived usefulness were slightly larger for JonDonym users. A possible explanation could be the structure of the two services, JonDonym is a profit-oriented company that charges for the unlimited use of the PET [110] while Tor is a community-driven project based on donations.

To gather some reasons for the observed differences and possibly identify other differences of the services from a user perspective, we included five open questions in the survey and collected altogether 626 statements, which we coded in two phases [33] with initial and focused coding. The results are shown in Tab. 4.3. In the left column, we have the three concepts technical issues, beliefs and perceptions and economical issues. Each of them includes several subconcepts. The results were then clustered into statements common to both PETs, such as feature requests (**Tor.1**, **Jon.1**), statements only referring to Tor, such as statements about malicious exit nodes (**Tor.2**), and statements only referring to Jondonym, such as concerns about the location of mix cascades (**Jon.2**). For each statement, we selected at least one quote shown at the bottom of the table.

The result indicates, that in the user perception both services differ not that much in technical issues but in the users' beliefs and perceptions. Unsurprisingly, economical issues were only concerning JonDonym. Three main differences might be able to explain the observed different effect sizes in the structural equation model. As already discussed, trust models between the services were different in the way that for JonDonym, users have to trust a company (**Jon.13**) while Tor users have to trust their community (**Tor.12**). While

Table 4.3: Results of the coding for the open questions including quotes

Concepts	Subconcepts	Common to both PETs	Specific Subconcepts for Tor	Specific Subconcepts for JD
Statements about Technical Issues	PET design	Feature Requests (**Tor.1**, **Jon.1**)	Malicious exit nodes (**Tor.2**)	Location of mix cascades (**Jon.2**)
	Compatibility	Accessibility of websites (**Tor.3**, **Jon.3**)		
	Usability	Documentation (**Tor.4**, **Jon.4**) Ease of use (**Tor.5**, **Jon.5**) Missing knowledge to use it correctly (**Tor.6**,**Jon.6**)		
	Performance	Latency (**Tor.7**, **Jon.7**, **Jon.8**)		
Beliefs and Perceptions	Anonymity	Concerns about deanonymization (**Tor.8**, **Jon.9**) Reason of use (**Tor.9**, **Jon.10**)		Size of the user base (**Jon.11**)
	Consequences	Fear of investigations (**Tor.10**, **Tor.11**, **Jon.12**)	Beliefs about social effects (**Tor.13**, **Tor.14**)	
	Trust		Trust in the community (**Tor.12**)	Trust in technology (**Jon.13**)
	Substitute technologies	Best available tool (**Tor.15**, **Jon.14**)		Tor as reference technology (**Jon.3**, **Jon.8**, **Jon.11**)
Statements about Economical Issues	Costs			Lower costs, other pricing schemes (**Jon.15**)
	Payment methods			Easy, anonymous payment options (**Jon.15**)
	Use cases		Circumvent Censorship (**Tor.16**)	Willingness to pay in certain scenarios (**Jon.16**, **Jon.17**)

Tor.1 *TCP support for name resolution via Tor's DNSPort [. . .]*
Tor.2 *Many exit nodes are run by governmental intelligence organisations. Exit notes can collect unencrypted data.*
Tor.3 *It can't be used on all websites; therefore it is of limited use to me*
Tor.4 *Easy to understand instructions for users with different levels of knowledge.*
Tor.5 *Tor protects privacy while on the web and is easy to use.*
Tor.6 *An unexperienced user may not understand the technical limitations of Tor and end up losing [. . .] privacy.*
Tor.7 *Increased latency makes the experience painful at times*
Tor.8 *It may fail to provide the expected level of anonymity because of attacks which may not even be known at the time they are performed (or commonplace).*
Tor.9 *It is a key component to maintaining one's privacy when browsing on the Internet.*
Tor.10 *Tor usage "Stands out"*
Tor.11 *[. . .] having a cop boot at my door because of Tor.*
Tor.12 *An end user needs to trust the network, the persons running Tor nodes and correct implementations [. . .]*
Tor.13 *Only social backlash from people thinking that Tor is mostly used for illegal activities.*
Tor.14 *For the same reason I don't hang out in brothels, using Tor makes you look like a criminal*
Tor.15 *While not perfect, Tor is the best option for reliable low-latency anonymization*
Tor.16 *It can be used as a proxy / VPN to get past censorship*

Jon.1 *Larger number of Mix Cascades, more recent software, i.e. preconfigured browser, faster security updates*
Jon.2 *First and last server of the mix cascade should not be located in the same country*
Jon.3 *Unlike Tor, JonDonym is not blocked by some websites. (Google for example among others)*
Jon.4 *Clearer explanations and instructions for JonDoFox*
Jon.5 *Easy to use, outside the mainstream like i.e. Tor*
Jon.6 *Privacy is less than expected because of wrong configuration settings.*
Jon.7 *[. . .] Even if it is quite slow without a premium tariff*
Jon.8 *[. . .] sometimes it's a little bit to slow, but compared with Tor...*
Jon.9 *Defeat of your systems by government agencies.*
Jon.10 *It provides a minimum level of personal data protection and online safety.*
Jon.11 *Tor is better due to having a much larger user base. More users results in greater anonymity*
Jon.12 *By using the service, am I automatically marked by intelligence authorities as a potential terrorist, supporter of terrorist organizations, user [. . .] for illegal things?*
Jon.13 *How can I trust Jondonym? How can Jondonym proof that servers are trustworthy?*
Jon.14 *It appeared to be the least worst option for anonymisation when I researched anonymisation services*
Jon.15 *Fair pricing, pre-paid is an easy payment option.*
Jon.16 *For use it in a country where it's difficult surf the net*
Jon.17 *If I would use the computer for work-related tasks*

the concept for both technologies is that the users' anonymity does not rely on a single malicious server, there is still trust necessary since only a minority of the users will inspect the programs they are running. For JonDonym users the size of the user base was also an issue (**Jon.11**). However, the most interesting observation also in terms of explaining the weak effect of behavioral intention on actual use behavior for Tor users was that many Tor users were concerned about looking like a criminal (**Tor.13**, **Tor.14**).

In summary, our results indicate that (with the newly introduced constructs perceived anonymity and trust in the PET) technology use models are applicable for PETs also. Most of the existing variables in the TAM

were also found in the participants' statements (e. g. usability, performance, anonymity and trust). However, our results can only be a first insight into issues of hindering a broader adoption of PETs, where more details have to be brought to light in future work.

4.1.2 Economic Incentives

Besides users' concerns and factors influencing their technology use acceptance, it is also important to consider factors for a successful business model built on a PET. For that purpose, we investigated the users' willingness to pay for a PET [89] and also considered the perspective of companies by investigating their incentives and hindrances to implement PETs [88].

Customers

Within the same survey as already described in the previous section, we also asked JonDonym users about their recent tariff and Tor users if the ever have donated to Tor [89]. It showed that the majority of users was not willing to pay or donate for the services: 85 out of 141 users (60%) used JonDonym's free tariff and 93 out of 124 (75%) Tor users have never donated to Tor.

For JonDonym, we also compared the users preferences for certain tariff structures depending on factors as data volume, pricing, and contract duration. We were comparing the users' preferences towards existing tariffs high-data-volume tariff, a low-price tariff, and a low-anonymity tariff and two newly created tariffs adding a lower data volume than the low-price-tariff and a higher volume than the high-data-volume tariff. Free users were neutral to all tariffs, but showed a slight preference to the newly created low-traffic tariff. Already paying users preferred the existing and newly created high-data-volume tariffs over the others. This indicates that free users would prefer the cheapest tariff if they decide to pay at all. This suggests that provider of PETs should offer tariffs with a low monetary barrier to convert free users to paying users. However, even with a low monetary barrier, there would still be the need to resolve the payment barrier, which regularly show in e-commerce when customers are abandoning their shopping cart before the payment process [176].

We also built a regression model to identify significant factors contributing to the willingness to pay. For that purpose, we defined a binary classifier for the willingness to pay (JonDonym), being 0 if the respondent was using a free tariff and being 1 if the respondent was using a premium tariff. Analogous, we defined the willingness to donate (Tor), being 0 if the respondent has never donated and being 1 if the respondent has donated at least once. As independent variables, we considered risk propensity (RP), frequency of improper invasion of privacy (VIC), trusting beliefs in online companies (TRUST), trusting beliefs in JonDonym (TRUST_{PET}) and knowing of Tor / JonDonym (TOR / JD) and derived the following research model:

$$WTP/WTD_i = \beta_0 + \beta_1 \cdot RP_i + \beta_2 \cdot VIC_i + \beta_3 \cdot TRUST_i + \beta_4 \cdot TRUST_{PET,i} + \beta_5 \cdot TOR/JD_i + \epsilon_i$$

The results are shown in Tab. 4.4 and one more time indicate that trust in the PET is the prevalent factor. On a highly significant level, the regression model suggests that a one unit increase in trust results in a roughly 12% higher likelihood that users choose a premium tariff (JonDonym) or donate (Tor). Besides that, there was only risk propensity significant for JonDonym and privacy victim for Tor. Surprisingly, risk propensity had a negative coefficient, indicating that more risk-averse users are less likely to choose a premium tariff for JonDonym. This contradicts previous findings [64] that risk aversion can act as a driver to protect an individual's privacy. For Tor, bad experiences with privacy breaches lead to a higher probability of donating money, even though on a more marginal level of roughly 5% per unit.

Companies

Equally important to the user perspective for the broad distribution of PETs is the perspective of the companies since user can only order services if they are offered. Therefore, we investigated the incentives and hindrances of companies to implement PETs either in their existing products or as a stand-alone product.

For that purpose, we conducted semi-structured interviews with 12 experts and managers from companies dealing with privacy and PETs in their daily business [88]. Our interview guide consisted of three relevant parts about general questions on the interviewees and their companies, technical questions on the status quo,

Table 4.4: Tor and Jondonym Users, Logistic Regression Model for Willingness to Donate/Pay [89]

Factor	WTP for JonDonym		WTD for Tor		Difference
	Coefficient	Avg. marg. effect	Coefficient	Avg. marg. effect	Avg. marg. effect
(Intercept)	-0.0376	-0.0081	6.1455***	-0.9768	0.9687
Risk Propensity	**-0.4967****	-0.1067	-0.1492	-0.0237	-0.083
Privacy Victim	-0.0397	-0.0085	**0.3352****	0.0533	-0.0618
Trust	-0.0868	-0.0187	-0.1222	-0.0194	0.0007
Trust_{PET}	**0.5661*****	0.1217	**0.7835*****	0.1245	-0.0028
Knowing Tor/Jondonym	-0.5792	-0.1245	0.488	0.0776	-0.2021

Significance: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

and questions on economic and societal issues. The interviews were recorded, transcribed, openly coded and in a second round selectively coded. The selective coding was done first separately and then among all interviews to consolidate the developed codings [33, 71]. We identified the following categories:

Technical Optimization PETs help to optimize the company within an organization and technical dimension and can get the company a technological lead. For that purpose the *integration into the business process* was named as a necessary condition and it was criticized that it is in general hard to get information about the practical use of PETs. PETs were also seen as a tool for *data management and avoidance* to improve business processes.

Business model The category considering business models was by far the largest. Here, the interviewees saw the largest incentives but also the largest hindrances. With the implementation of PETS, companies intend to *further development of services*. How and if that works, depends on the customers' requirements, if the level of convenience for the existing service (if it depends on customer data) as well as of the PET's handling. Customers' awareness of privacy was also seen as an important factor. However, the interviewees were discordant if raising it should be the task of the company. PETs were also seen as a chance to *enlarge the company's clientele* by addressing nerds. The mass market was seen from the viewpoint that most customers do not request PETs, but would accept them and that it offers a chance to implement PETs in existing products which are already widespread. Interviewees did also not agree on the *development of new business models* in terms of offering privacy as a premium feature. While some considered it as naturally to ask for a fee for the additional effort on the company's side, others questioned that approach by referring to the perception of the 'non-premium' customers that they do not have a sufficient security and privacy when using the company's service. As a last incentive a better *positioning for the future* was named which could gain the company an advantage over its competitors.

Corporate perception The particular technology was considered to be less important, but a positive perception by business partners was considered to be highly useful to gain *trust*. Using PETs to have a communicable unique selling point enables the company to *profile itself through PETs*. *Business ethics* was considered from multiple viewpoints by considering anonymity as neutral technology, using the customer's fear to sell them PETs or integrating PETs because it seems to be the right thing to do.

Our results do not draw a clear picture in some areas since the perceptions differ a lot, i. e. on the question if privacy can be sold to the customers as a premium service. This shows that more research is necessary to determine underlying factors and elaborate precise recommendations to companies how they can integrate PETs in their products while having a proper business model in mind.

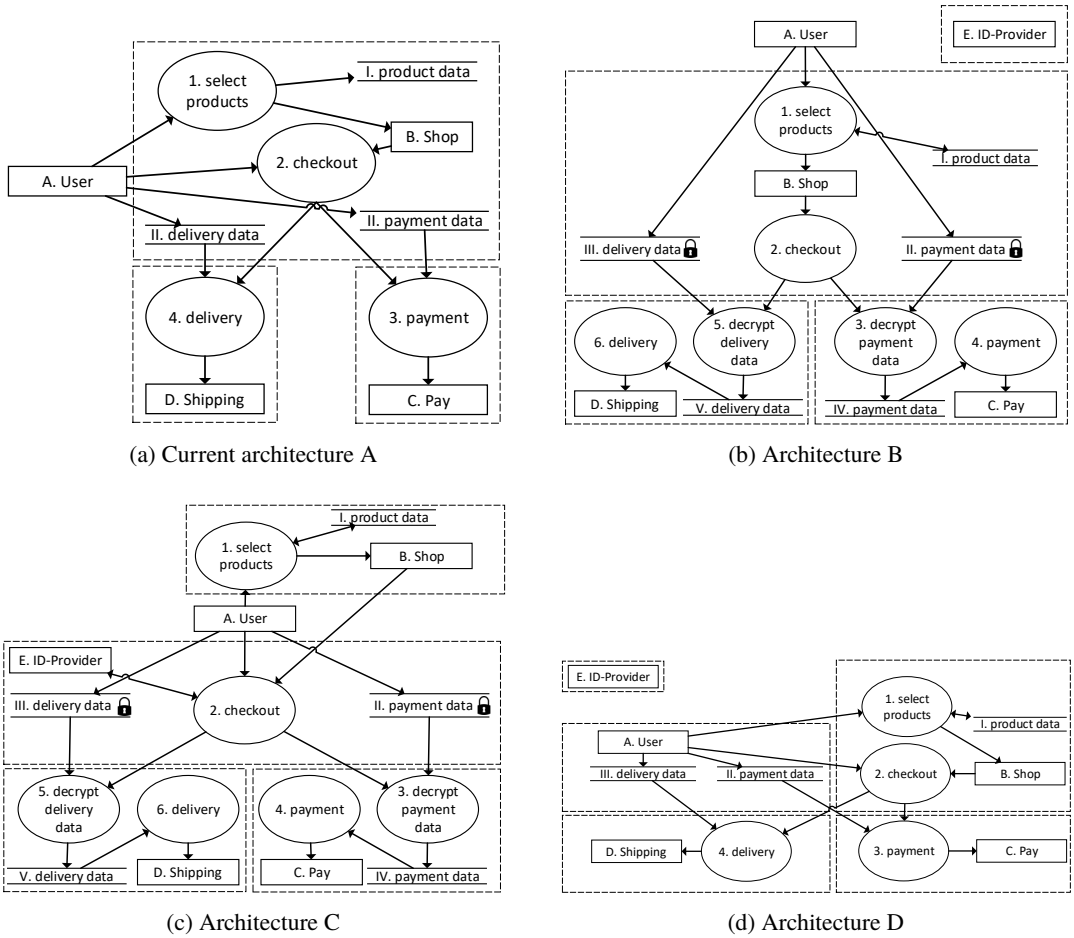
4.2 Privacy by Design

In this section, we demonstrate how existing approaches can be used and apply privacy by design in practice. For that purpose, we consider two use cases: Online shopping and the internet of things, in particular for an

autonomous driving scenario. In both approaches, existing technology and patterns are used to improve the users' privacy.

4.2.1 E-Commerce

The importance of online shopping has steadily increased during the last years. Despite an increase in public's awareness on the issue of data protection and growing concerns about the usage of their data, currently users of online shopping platforms have no alternative to disclosing personal data. With the General Data Protection Regulation (GDPR) being in place and asking to implement data-protection principles, such as data minimization, the aim was to improve the processes in e-commerce with respect to the data collected by involved parties. In 2016, a study revealed that 50% of online services send full information about the users' baskets to Paypal if PayPal was selected as payment method [170]. Even though online shopping platforms could track users by technical measures such as browser fingerprinting [60] or evercookies [4]. As previous work has already suggested different countermeasures [14, 55, 122, 167], we focused on the information flow for the basic online shopping processes. For that purpose, we suggested four different architectures for building such a platform as shown in Fig. 4.5. The intend was to minimize data across all parties [157], in particular the online-shop does not need to learn the identity of the user, a payment provider, and delivery service only need to learn payment or delivery information respectively, but both of them do not need to know details about the purchase. This way, the user could shop pseudonymously, but none of the involved parties learns who bought what.



(a) Current architecture A

(b) Architecture B

(c) Architecture C

(d) Architecture D

Figure 4.5: Data Flow Diagram for Different Architectures in E-Commerce [157]

Architecture A (cf. Sect. 4.5a) shows the current status quo where the payment and delivery data is sent via the online shop to the payment provider and the delivery services. In Architecture B (cf. Sect. 4.5b), this data is also sent via the shop, but the user encrypts it beforehand with the public keys of the payment provider and the delivery services, therefore, the online shop doesn't learn it. An identity provider knows the identity of the user and provides a login service. Thus, the shop does not know the identity of the user, but the identify provider could reveal it in case a problem arises. Architecture C (cf. Sect. 4.5c) is very similar to Architecture B, the only difference is that the encrypted payment and delivery data are stored at the identity provider. In Architecture D (cf. Sect. 4.5d), the user gets redirected by the shop to the payment and delivery service and directly provides the payment and delivery data to them.

We then compared the different architectures based on the privacy threat analysis methodology LINDDUN [217] as shown in Tab. 4.5, but also with respect to usability, transparency and compatibility to existing business models. For that purpose, we assumed that the different parties were not colluding to profile users. The considered threats were identifying the user, learning the content of the shopping cart, the total value of

Table 4.5: Privacy Threats Mapped to Architecture Variants in E-Commerce [157]

Threat \ Entity	Shop	Pay	Ship	Identity Provider
Identifiability	A	(ABCD)[2]	(ABCD)[3]	B C D
Disclosure shopping cart	A B C D			
Disclosure total value	A B C D	A B C D		
Disclosure payment data	A	A B C D		
Disclosure delivery data	A		A B C D	
Linkability purchase	A(BCD)[1]	(ABCD)[2]	(ABCD)[3]	C
Detectablility login	A B C D			B C D
Detectablility purchase	A B C D	A B C D	A B C D	C
Detectablility payment	A B C D	A B C D		C
Detectablility delivery	A B C D		A B C D	C

[1] Depends on the user's choice. [2] Depends on user's payment [3] Depends on user's shipping

the purchase, the payment data, and the delivery data. Also considered was if different purchases can be linked which can be desirable by the user to get a reputation as a good customer or avoided for privacy reasons. Furthermore, we considered which parts of the shopping process (login, purchase, payment, delivery) which of the involved parties was able to detect (for a non-interrupted shopping process).

For the overall evaluation two criteria are decisive: privacy and usability. *Privacy*: all proposed architectures only differ in the amount of information the identity provider is able to learn. *Usability*: Architecture B and C require some effort since the users have to encrypt their data, but would allow to delegate that task to the shop or the identity provider. However, delegating it to the shop would foil the whole approach. Architecture D requires the user to provide his payment and delivery data for each purchase again. Furthermore, since the shop needs to redirect the user to the payment provider or the delivery service, the user has to check the identity of them in each purchase also.

For the final evaluation, even though architectures B and D are favorable over architecture C in terms of privacy, the advantage of architecture C in terms of usability is decisive since it is the only architecture where it is reasonable for the users to delegate the encryption and provision of their data. Therefore, we believe architecture C to be the most feasible option for now. This might change in future to architecture D in case browsers are able to take over this task, e. g. by a widespread implementation of the PaymentRequest API [13].

4.2.2 Internet of Things

The internet of things (IoT) can be seen as a large network of connected sensors. Therefore, security and privacy have a decisive role. While the security problems for IoT devices, such as missing updates or difficulties for users to install patches without keyboard and display, are far from solved, there are promising

ideas to raise the companies' incentives to not neglect previous buyers. Morgner et al. [134] consider regulatory approaches to request mandatory security update labels by the companies that inform consumers during buying decisions about the willingness of the manufacturer to provide security updates in the future. For this section, we focus on the privacy of the internet of things. Studies indicate that 60% of the IoT devices don't properly tell customers how their personal information is being used [100] and almost all IoT areas miss applicable mechanisms in privacy [129]. Therefore, we investigated privacy policies for IoT devices [165] and found that most of them were neither transparent not privacy friendly confirming the findings of Information Commissioner's Office [100]. Additionally, we show how to apply existing privacy patterns to the IoT by making use of its underlying architecture to improve users' privacy [154].

Privacy Policies

To investigate IoT privacy policies, we developed a framework to asses IoT privacy policies based on the GDPR [165]. Since the GDPR "lays down rules relating to the protection of natural persons with regard to the processing of personal data" [61, Article 1 para. 1], it is also addressed to suppliers of IoT products.

The developed framework consists of 16 parameters (as shown in Tab. 4.6) with all besides the first of them having up to four "yes-or-no" questions. We identified two important dimensions for the framework: (i) Content-Dimension (Privacy Score) and (ii) Transparency-Dimension (Transparency Score). They differ in so far that the transparency-dimension rather checks whether the policy makes a statement or not and the content-dimension rather checks what statement the policy makes. For all categories, we did a legal assessment to check how we should cope with a nonexistent statements. Some of these statements are mandatory, e. g. mentioning the user's right to object, and therefore their absence is considered negative. Other statements are optional statements, e. g. statements about sharing the data with third parties. Their absence in theory means that the user's data is not shared with third parties, and therefore their absence is considered positive (for the privacy score). The different categories are weighted to either have all the same score or users can give their priorities for a weighted scoring of the different categories.

We manually assessed privacy policies of 110 different IoT devices for which we were able to find English privacy policies. Table 4.7 provides an overview of the outcome. For a better overview, we have classified the devices into different areas and subcategories. However, the scores seemed to be only marginally different, so we did not further investigate differences between the individual groups.

The results of our assessment showed that most of the examined privacy policies of IoT devices/services were insufficient to address the GDPR requirements and beyond. However, since the assessment was done before May 2018, companies had some time to address the issues before the GDPR came into effect.

Privacy Patterns

The IoT heavily relies on a cloud and fog computing infrastructure as shown in Fig. 4.6. The initial idea of cloud computing was to have central large data centers with lots of computational resources. However, it has shown that this architecture is not optimal if a minimal latency is required. Therefore, the architecture was extended by two more layers between the cloud computing center and the user denoted as fog and edge computing [107]. So far, the fog and edge computing architecture had been used to reduce latency, improve contextual location awareness or scalability, but not to address the users' privacy. On the other hand, while some approaches to address privacy issues in the IoT existed, none of the made use of the specific underlying architecture [219]. To avoid re-inventing the wheel, we applied privacy patterns [57, 73] to this infrastructure to demonstrate how the underlying architecture may be used for privacy purposes [154]. Similar to software design patterns, privacy patterns are solutions to common privacy issues, which can be used as guidance. For each of the patterns we also provided an example from smart vehicle scenarios to demonstrate their application.

Figure 4.7 shows two examples out of the seven patterns provided in the paper: The personal data store and data isolation at different entities.

The personal data store pattern (cf. Fig. 4.7a) suggest that users keep control about their personal data and store it on a personal device, e. g. their mobile phone. This pattern is most useful for data produced by the user. As a consequence, if possible computations on that data can also be done locally. If the IoT device is too small and has too few computational power, a workaround would be to make use of the user's mobile

Table 4.6: Parameters for the Framework to Assess Privacy Policies [165]

#	Parameter Name	Parameter Description	T	P	§
1	Easily Acc. Form	1) Readability (Flesch Reading Ease Score)	✓	✓	12
2	Right to Object	1) Does the policy state a right to object? 2) Is an objection as easy as a consent?	✓	🔍	6, 7, 13, 21
3	Children	1) Is a binding age limit to use the service stated? 2) Is there a special policy for children? 3) Is there a mechanism to ensure that parents agree with the processing? 4) Does the policy state the procedure if children data has been processed unintentionally?	✓	🧍	8
4	Processing of Special Categories of Personal Data	1) Are special personal data categories processed? 2) Is it required contentwise for using the service? 3) Is there an explicit consent?	✓	🔍+	9, 13
5	Necessary Information	1) Are identity and contact details of the controller stated? 2) Is a data protection officer stated? 3) Are the purposes of the processing for which the personal data are intended stated?	✓	🔍	13
6	Period of Storage	1) Is the storage period stated? 2) Are criteria determining the period stated?	✓	🔍	13
7	Right of Access	1) Is the right of access stated? 2) Is a fee charged?	✓	🔍	12, 13, 15
8	Right to Erasure	1) Is the right to erasure stated? 2) Is the time to fulfil the erasure request stated? 3) Period until fulfilment	✓	🔍	12, 13, 17
9	Data Portability	1) Is the right to data portability mentioned?	✓	🔍	13, 20
10	Third Countries	1) Is data processed in third countries? 2) Does the policy state these countries? 3) Is data transferred to countries with adequate level of protection (e.g. EU-U.S. Privacy shield)?	✓	🔍+	45, 46, 47, 49
11	Data Breach Notification	1) Is a personal notification after a data breach explicitly stated? 2) Period until notification	✓		34
12	Third Parties	1) Is a third party involved by design? 2) Does the policy state who the third party is? 3) Does the policy explicitly state the purpose? 4) Is the scope of the transferred data stated?	✓	🔍+	13
13	Search for the Policy	1) Is there a link on the homepage that leads to the policy for the device quickly? 2) How many clicks are needed from the homepage to find the link to the policy?	✓		12, 13
14	Change Notificat.	1) Is there a notification after policy changes?	✓		13
15	Special Device Policy	1) Is the present policy a multi-policy? 2) Is it clear, the policy is for the IoT product?	✓	✓	
16	Lifecycle	1) Can information stored on the device be deleted?	✓	🔍	

✓: Used, : Not used, 🔍+/🔍: If not present, rated positive/negative, 🧍: Only for toys
T: Transparency, P: Privacy Friendliness of the Policy

Table 4.7: Summary Statistics of Examined Policies [165]

Area	Subarea	#	PPS Score A	B	C	D	E	Rel. PPS (%) Mean	STD	Transparency A	B	C	D	E	Rel. TS (%) Mean	STD
Smart Home	Coffee Machine	5	0	0	1	4	0	31.67	8.39	0	0	4	1	0	47.50	10.37
	Light	5	0	0	2	3	0	35.56	8.67	0	1	4	0	0	53.75	6.04
	Security	9	0	0	3	5	1	32.80	11.36	0	1	7	1	0	48.61	9.80
	Thermostat	6	0	0	3	3	0	36.69	11.10	0	1	4	1	0	50.43	11.35
	Washer	5	0	1	2	2	0	37.91	20.83	0	1	3	1	0	54.17	12.68
	Others	28	0	0	7	21	0	34.71	8.95	0	5	20	3	0	50.52	8.99
	Total	58	0	1	17	38	2	34.70	10.50	0	9	42	7	0	50.55	9.37
Health	Fitness Tracker	7	0	0	2	5	0	36.11	6.39	0	1	6	0	0	53.72	4.91
	Scale	15	0	0	1	12	2	28.75	11.56	0	3	6	6	0	43.89	12.93
	Others	5	0	0	1	4	0	33.89	8.22	0	1	4	0	0	52.29	6.93
	Total	27	0	0	4	21	2	31.61	10.14	0	5	16	6	1	47.99	11.18
	Toy	9	0	0	3	6	0	34.05	12.66	0	2	6	1	0	50.92	13.18
∑	Total	94	0	1	24	65	4	33.75	10.59	0	16	64	14	0	49.85	10.26

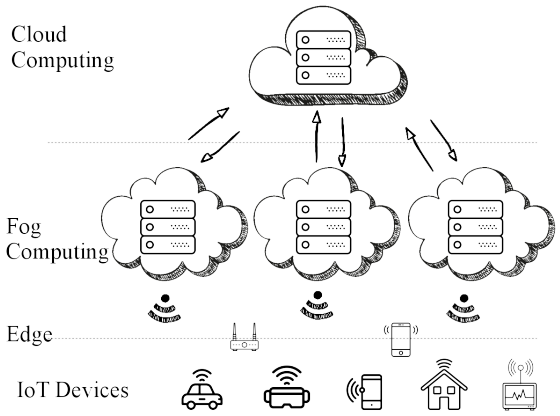
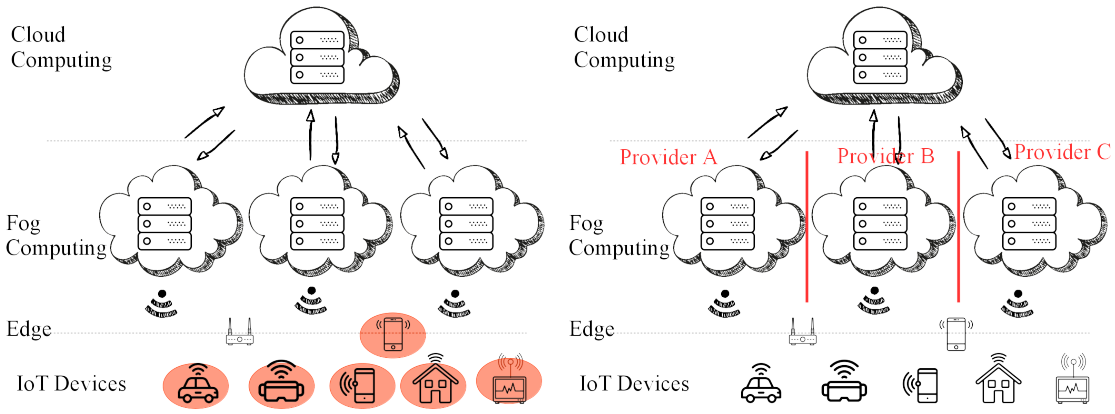


Figure 4.6: Three-layer service delivery model [154]



(a) Personal Data Store

(b) Data Isolation at Different Entities

Figure 4.7: Privacy Patterns Applied to the IoT / Cloud Computing / Fog Computing Architecture [154]

phone. Many devices connect to the internet via the user's phone anyway or use the user's phone with an app

as interface to control the device. However, nowadays often the data is stored in the cloud and accessed by the phone's app, even though this is not necessary.

The data isolation at different entities pattern suggest that data or usage information is distributed among several entities, so that each entity may only see a part of the data. In the determined architecture, the fog nodes or clusters would be an excellent layer to enforce isolation (cf. Fig 4.7b). In particular, if fog nodes are clustered vertically, each cluster could belong to a different organization. Given that fog clusters of an organization would be only at one location, this would improve the users' privacy by preventing the companies to build global profiles on them.

By applying existing privacy patterns to the IoT architecture we could demonstrate how the specific properties of this architecture can also be used improve the users' privacy. However, if a pattern can be applied, it needs to be investigated for each use case individually since it will also depend on its influence on the other desired fog computing properties, i. e. low latency.

Chapter 5

Discussion and Conclusion

> I urge you to bear in mind the imperfection of our current knowledge. Science is never finished. It proceeds by successive approximations, edging closer and closer to a complete and accurate understanding of nature, but it is never fully there.
>
> *Carl Sagan*

The previous three sections dealt with requirement elicitation and tool-support for social engineering, security management and privacy-enhancing technologies. Although the three areas are fundamentally different, they share a lot of commonalities. With humans involved in all areas, interdisciplinary is the most preeminent one. However, it is worth to have a closer look at the other commonalities, also.

Awareness and Serious Games Awareness is a reoccurring topic for fighting social engineering and the use of privacy enhancing technologies. In particular for privacy, users have shown lack of knowledge and unawareness of the consequences resulting from privacy leaks [69]. As a first strategy, it seems to be important to build a security and privacy culture supporting users in their efforts, so that they are at least aware and can count on support from others. Furthermore, the use of serious games or the integration of educational content in other everyday media, such as television series, could raise the users' awareness without the need to participate in training and awareness raising courses.

Convenience Users and companies have a certain degree of convenience in common. Users do not make use of privacy enhancing technologies, because their use is too complicated and/or takes too much effort [69]. Earlier research found also that it is important to "understand the target population" and research suggesting zero-effort privacy [87, 97] by improving the usability of the service and removing obstacles to reduce the user's necessary effort. We could notice a similar behavior when we investigated energy providers. Most of them did only the minimum necessary to fulfill the legal obligations. Their sketched desire was to buy a box for security, which will just be integrated into their network and does not need any other maintenance.

Decision Making In all three areas certain outcomes depend on the users' or managers' decision making on imprecise information (cf. also next paragraph). Users often are insecure if they are attacked by a social engineer. If they are more confident, that they can handle the situation well (attitude bolstering), their outcome improves and they are able to fend off more attacks. Managers do hardly ever get feedback if they decided to spend "too much" on security. Most times it's only vice versa, if they spent to little, this might result in successful attacks or even failures in compliance. Also for privacy, it's hard for the users to make the correct decisions, e. g. many users do not seem to be aware of specific privacy risks when confronted with an abstract risk scenario [68].

Difficulties to Measure Security and Privacy One of the difficulties of security management is the difficulty to measure security directly [26]. The management is used to take decisions on the basis of key

performance indicators. However, for security it can be very hard to create some meaningful metrics. As the defense against social engineering attacks is also part of the security management of a company, this concerns social engineering defenses as well. One could try to measure the company's resistance by using social engineering penetration testing (including the simulation of phishing attacks [210]). However, this could influence the employees' motivation and would need clearance from the works council. Therefore, it is also reasonable hard to assess the effects of training and awareness raising in practice.

In a similar manner, it is difficult to assess the impact of privacy enhancing technologies. While it might be manageable to measure or guarantee certain privacy metrics such as k-anonymity [204] or differential privacy [58], it is almost impossible for the users of a privacy enhancing technology to asses how and in which way they have been protected from profiling, data leakage, etc.

Economic Incentives Economic incentives come into play in three variations: First, for security management, there is always the trade-off between investments in security and the possible damage resulting of a successful attack which could have been defended. This also includes questions about the optimal security strategy in terms of resources for a given budget. Second, when considering cyber criminals, they also have an economic perspective regarding their attacks. Since they want to make money, they will not spend more resources on an attack than they assume to get a return on investment. Third, for the dissemination of privacy enhancing technologies, it is important that companies privacy enhancing technologies somehow support the companies' business model or can build a business model upon privacy enhancing technologies.

Regulations and Legal Aspects It is obvious that regulations play a major role in security management due to its proximity to compliance. With the implementation of the GDPR [61] – including a maximum fine of 4% of the company's annual global turnover for privacy violations – this became similar for privacy enhancing technologies and privacy regulations became part of the compliance requirements in companies. One can observe two consequences. On the one hand, the regulations can act as a driver, e. g. for implementing an information security management system or privacy enhancing technologies. On the other hand, many companies act according to the motto "a clever horse doesn't jump higher than necessary". Thus, even if they see possibilities to easily improve their security or privacy, they don't do it as long as they fulfill the compliance requirements.

For social engineering it was not that obvious that regulations play an important role. However, besides the requirement of some standards (e. g. ISO 27001 [104]) to train employees, the task of defending against social engineering is closely connected to labor law. Many of the possible measures, such as penetration testing, are restricted or need clearance by the works council. This also concerns serious games if the personality of the employees is part of the game.

Note that many of the described commonalities also interfere with each other. When looking at privacy enhancing technologies for example, lack of awareness, convenience of users and difficulties in measuring the impact of privacy enhancing technologies negatively influence the users willingness to pay and therefore the economic incentives for a company to offer privacy enhancing technologies vanishes. However, economic impact could be overruled by regulations requiring certain service providers to implement privacy enhancing technologies.

Future Work

Since science is never finished, and often an answered question raises even more new questions, this section lists proposals for future work for already ongoing work building on top of the presented contents.

Social Engineering

A survey investigating serious games on information security and providing a structured overview is currently work in progress. The continuous evaluation and improvement of the games HATCH, PROTECT and the

CyberSecurity Awareness Quiz is also ongoing, along with considerations how to complement the concepts with additional games and the integration into security training platforms.

A particular interesting question concerns gender stereotypes in serious games, and in particular HATCH. Hill et al. [98] showed that the use of multiple photos (of males and females) for a single persona to avoid gender stereotypes did not reduce project designers' engagement with the personas in requirement engineering. However, it is unclear if the same holds in a serious game and the use of multiple photos for a single persona would change the players' engagement with HATCH's personas.

Security Management

We have already discussed commonalities regarding the regulations concerning information security and privacy management. These have already been taken up by the technical committee ISO/IEC JTC 1/SC 27 and resulted in the standard ISO/IEC 27701 [106] which is a privacy extension to ISO/IEC 27001. This way, the information security management system is enhanced to include a privacy information management system allowing a holistic management for security and privacy. It would be interesting to further adapt and evaluate the presented tools and frameworks to cover also the extension.

Regarding the selection of a secure cloud service provider, several ideas come to mind. Naturally, an implementation on a website would allow to evaluate the concept in practice. To extent the selection process to cover more relevant criteria, performance, costs, and the costs to migrate to another provider could be added to the comparison making security only one of the considered criteria. Furthermore, so far our proposed approach is not feasible to consider further advanced architectures such as the Cloud-of-clouds [22] or other work proposing changed trust assumptions by making use of tamper-proof hardware containers and third party audits [2]. It would also be interesting to extend the developed cloud battlefield model to include more complex architectures.

Privacy Enhancing-Technologies

Regarding the users' privacy concerns and adoption of privacy enhancing technologies, it would be interesting to also investigate the users' knowledge about and the influence of malicious tor relays which are exploiting users [140].

Further work in progress also regards to sensors in the internet of things. By developing a framework to assess the risks associated to different sensors and in particular inference attacks on them (cf. Sikder et al. [199]), it would be possible to allow the users an informed decision on privacy risks associated with the internet of things devices they use. This could result in a privacy label similar to security labels for internet of things devices proposed by Morgner et al. [134]. Since many devices are controlled by apps anyways, the idea could also be integrated in existing approaches informing the user about privacy risks for apps as proposed by Hatamian and Serna-Olvera [91].

Another relevant question regarding privacy enhancing technologies concerns the storage of data either at the users' devices or within a cloud computing environment. From a privacy perspective, it seems to be preferable to store personal data at the users devices. However, given that internet of things devices and also smartphones lack updates because manufacturers are not providing them at all or stopping support after a certain time, from a security perspective that may not be the best decision. The problem becomes even more challenging if nodes in between, such as fog computing nodes are also considered and additionally offer the distribution of data across several entities. Are there underlying factors in which use case for which data which of the different approaches might be preferable?

Bibliography

[1] Rana Khudhair Abbas Ahmed. Security Metrics and the Risks: An Overview. *International Journal of Computer Trends and Technology*, 41(2):106–112, 2016. ISSN 22312803.

[2] Lamya Abdullah, Felix C. Freiling, Juan Quintero, and Zinaida Benenson. Sealed computation: Abstract requirements for mechanisms to support trustworthy cloud computing. In *Computer Security - ESORICS 2018 International Workshops, CyberICPS 2018 and SECPRE 2018, Barcelona, Spain, September 6-7, 2018, Revised Selected Papers*, pages 137–152, 2018. doi: 10.1007/978-3-030-12786-2_9. URL https://doi.org/10.1007/978-3-030-12786-2_9.

[3] Ruba Abu-salma, M. Angela Sasse, Joseph Bonneau, Anastasia Danilova, Alena Naiakshina, and Matthew Smith. Obstacles to the Adoption of Secure Communication Tools. In *IEEE Security & Privacy*, pages 137 – 153, 2017. doi: 10.1109/SP.2017.65.

[4] Gunes Acar, Christian Eubank, Steven Englehardt, Marc Juarez, Arvind Narayanan, and Claudia Diaz. The web never forgets: Persistent tracking mechanisms in the wild. In *CCS*, 2014.

[5] Icek Ajzen. The Theory of Planned Behavior. *Organizational Behavior and Human Decision Processes*, 50(2):179–211, 1991. ISSN 07495978. doi: 10.1016/0749-5978(91)90020-T.

[6] George A. Akerlof. The market for 'lemons': Quality uncertainty and the market mechanism. *Quarterly Journal of Economics*, 84(3):488 – 500, 1970.

[7] Dina Aladawy, Kristian Beckers, and Sebastian Pape. PERSUADED: Fighting Social Engineering Attacks with a Serious Game. In Steven Furnell, Haralambos Mouratidis, and Günther Pernul, editors, *Trust, Privacy and Security in Digital Business - 15th International Conference, TrustBus 2018, Regensburg, Germany, September 5-6, 2018, Proceedings*, volume 11033 of *Lecture Notes in Computer Science*. Springer, 2018. ISBN 978-3-319-98384-4. doi: 10.1007/978-3-319-98385-1_8. URL https://doi.org/10.1007/978-3-319-98385-1_8.

[8] Michael Alexander. Methods for understanding and reducing social engineering attacks. *SANS Inst.*, 1: 1–32, 2016. URL https://www.sans.org/reading-room/whitepapers/critical/methods-understand-ing-reducing-social-engineering-attacks-36972.

[9] Ahmed M. Azab, Peng Ning, Emre C. Sezer, and Xiaolan Zhang. HIMA: A Hypervisor-Based Integrity Measurement Agent. In *Proceedings of the 2009 Annual Computer Security Applications Conference*, ACSAC '09, pages 461–470, Washington, DC, USA, 2009. IEEE Computer Society. ISBN 978-0-7695-3919-5. doi: 10.1109/ACSAC.2009.50.

[10] Ahmed M. Azab, Peng Ning, and Xiaolan Zhang. SICE: a hardware-level strongly isolated computing environment for x86 multi-core platforms. In *Proceedings of the 18th ACM conference on Computer and communications security*, CCS '11, pages 375–388, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0948-6. doi: 10.1145/2046707.2046752.

[11] Maria Bada, Angela M. Sasse, and Jason R. C. Nurse. Cyber security awareness campaigns: Why do they fail to change behaviour? *CoRR*, abs/1901.02672, 2019. URL http://arxiv.org/abs/1901.02672.

[12] Gabriel Bassett, C. David Hylender, Philippe Langlois, Alexandre Pinto, and Suzanne Widup. Data breach investigations report, 2020. URL https://enterprise.verizon.com/resources/reports/2020-data-breach-investigations-report.pdf.

[13] Adrian Bateman, Zach Koch, Roy McElmurry, Domenic Denicola, and Marcos Cáceres. Payment request api. https://www.w3.org/TR/2017/CR-payment-request-20170921/, 2017. W3C Candidate Recommendation 21 September 2017.

[14] Jason Bau, Jonathan Mayer, Hristo Paskov, and John C Mitchell. A promising direction for web tracking countermeasures. *W2SP*, 2013.

[15] BBC News. Nigerian men arrested over German PPE 'scam', Sep 2020. URL https://www.bbc.com/news/world-africa-54051424.

[16] Kristian Beckers and Sebastian Pape. A serious game for eliciting social engineering security requirements. In *Proceedings of the 24th IEEE International Conference on Requirements Engineering*, RE '16. IEEE Computer Society, 2016. doi: 10.1109/RE.2016.39.

[17] Kristian Beckers, Sebastian Pape, and Veronika Fries. HATCH: Hack and trick capricious humans – a serious game on social engineering. In *Proceedings of the 2016 British HCI Conference, Bournemouth, United Kingdom, July 11-15, 2016*, 2016. URL https://ewic.bcs.org/content/ConWebDoc/56973.

[18] Kristian Beckers, Veronika Fries, Eduard C. Groen, and Sebastian Pape. Creativity techniques for social engineering threat elicitation: A controlled experiment. In *Joint Proceedings of REFSQ-2017 Workshops, Doctoral Symposium, Research Method Track, and Poster Track co-located with the 22nd International Conference on Requirements Engineering: Foundation for Software Quality (REFSQ 2017), Essen, Germany, February 27, 2017*, volume 1796, 2017. URL https://ceur-ws.org/Vol-1796/creare-paper-1.pdf.

[19] Kristian Beckers, Daniel Schosser, Sebastian Pape, and Peter Schaab. A structured comparison of social engineering intelligence gathering tools. In *Trust, Privacy and Security in Digital Business - 14th International Conference, TrustBus 2017, Lyon, France, August 30-31, 2017, Proceedings*, pages 232–246, 2017. doi: 10.1007/978-3-319-64483-7_15. URL https://doi.org/10.1007/978-3-319-64483-7_15.

[20] Mathieu Bédard. The underestimated economic benefits of the internet. Regulation series, The Montreal Economic Institute, 2016. Economic Notes.

[21] Michel Benaroch. Real options models for proactive uncertainty-reducing mitigations and applications in cybersecurity investment decision making. *Information Systems Research*, 29(2):315–340, 2018. doi: 10.1287/isre.2017.0714. URL https://doi.org/10.1287/isre.2017.0714.

[22] Alysson Bessani, Miguel Correia, Bruno Quaresma, Fernando André, and Paulo Sousa. Depsky: dependable and secure storage in a cloud-of-clouds. *ACM Transactions on storage (TOS)*, 9(4):1–33, 2013.

[23] Martin Bland. Estimating Mean and Standard Deviation from the Sample Size, Three Quartiles, Minimum, and Maximum. *International Journal of Statistics in Medical Research*, 2015.

[24] Sören Bleikertz, Toni Mastelic, Sebastian Pape, Wolter Pieters, and Trajce Dimkov. Defining the cloud battlefield – supporting security assessments by cloud customers. In *Proceedings of IEEE International Conference on Cloud Engineering (IC2E)*, pages 78–87, 2013. doi: 10.1109/IC2E.2013.31.

[25] Rainer Böhme. Security metrics and security investment models. In *Advances in Information and Computer Security - 5th International Workshop on Security, IWSEC 2010, Kobe, Japan, November 22-24, 2010. Proceedings*, pages 10–24, 2010. doi: 10.1007/978-3-642-16825-3_2. URL http://dx.doi.org/10.1007/978-3-642-16825-3_2.

[26] Rainer Böhme. Security metrics and security investment models. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 6434 LNCS, pages 10–24, 2010. ISBN 3642168248.

[27] Bill Briggs, Kristi Lamar, Khalid Kark, and Anjali Shaikh. Manifesting legacy: Looking beyond the digital era. 2018 global CIO survey. Technical report, Deloitte, 2018.

[28] Bundesamt für Sicherheit in der Informationstechnik (BSI). KRITIS-Sektorstudie Energie. Technical report, Bundesamt für Sicherheit in der Informationstechnik (BSI), 02 2015.

[29] Bundesministerium für Arbeit und Soziales. Verordnung über Arbeitsstätten (Arbeitsstättenverordnung - ArbStättV) Arbeitsstättenverordnung vom 12. August 2004 (BGBl. I S. 2179), zuletzt durch Artikel 1 der Verordnung vom 30. November 2016 (BGBl. I S. 2681) geändert, 2016. URL https://www.bmas.de/SharedDocs/Downloads/DE/PDF-Publikationen/A225-arbeitsstaettenverordnung.pdf.

[30] Bundesnetzagentur. About us, September 2013. URL https://www.bundesnetzagentur.de/EN/Areas/Energy/AboutUs/aboutus-node.html.

[31] Bundesnetzagentur. Listen der Netzbetreiber und Versorgungsunternehmen, December 2019. URL https://www.bundesnetzagentur.de/DE/Sachgebiete/ElektrizitaetundGas/Unternehmen_Institutionen/HandelundVertrieb/Lieferantenanzeige/lieferantenanzeige-node.html.

[32] Mahinthan Chandramohan and Hee Beng Kuan Tan. Detection of mobile malware in the wild. *Computer*, 45(9):65–71, Sept 2012. ISSN 0018-9162. doi: 10.1109/MC.2012.36.

[33] Kathy Charmaz. *Constructing Grounded Theory*. Sage Publications, London, 2nd edition, 2014. ISBN 9780761973522. doi: 10.1016/j.lisr.2007.11.003.

[34] Anubhav Chitrey, Dharmendra Singh, and Vrijendra Singh. A comprehensive study of social engineering based attacks in india to develop a conceptual model. *International Journal of Information and Network Security (IJINS)*, 1(2):45–53, 2012.

[35] Jessica Clement. Number of internet users worldwide 2005–2019. *Statista Research*, 2020. URL https://www.statista.com/statistics/273018/number-of-internet-users-worldwide/.

[36] Cloud Security Alliance. The notorious nine: Cloud computing top threats in 2013. Technical report, Cloud Security Alliance, 2013. URL https://cloudsecurityalliance.org/download/artifacts/the-notorious-nine-cloud-computing-top-threats-in-2013/.

[37] Cloud Security Alliance. The treacherous 12 - cloud computing top threats in 2016. Technical report, Cloud Security Alliance, 2016. URL https://downloads.cloudsecurityalliance.org/assets/research/top-threats/Treacherous-12_Cloud-Computing_Top-Threats.pdf.

[38] Cloud Security Alliance. Consensus assessments initiative questionnaire. https://cloudsecurityalliance.org/artifacts/consensus-assessments-initiative-questionnaire-v3-1/, November 2019. v3.0.1.

[39] Cloud Security Alliance. Cloud controls matrix. https://cloudsecurityalliance.org/research/cloud-controls-matrix/, March 2019. v3.0.1.

[40] Cloud Security Alliance. Top threats to cloud computing the egregious 11. Technical report, Cloud Security Alliance, 2019. URL https://cloudsecurityalliance.org/download/artifacts/top-threats-to-cloud-computing-egregious-eleven/.

[41] Cloud Security Alliance. Homepage. https://cloudsecurityalliance.org/, 2020.

[42] Juliet Corbin and Anselm Strauss. *Basics of qualitative research: Techniques and procedures for developing grounded theory*. Sage publications, 2014.

[43] Pedro Costa, João Carlos Lourenço, and Migueli Mira da Silva. Evaluating cloud services using a multiple criteria decision analysis approach. In *Service-Oriented Computing*, volume 8274, pages 456–464. Springer Berlin Heidelberg, 2013.

[44] Fred D. Davis. A Technology Acceptance Model for Empirically Testing New End-User Information Systems: Theory and Results. *Massachusetts Institute of Technology*, 1985.

[45] Fred D. Davis. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly*, 13(3):319–340, 1989.

[46] Julian Dax, Benedikt Ley, Sebastian Pape, Christopher Schmitz, Volkmar Pipek, and Kai Rannenberg. Elicitation of requirements for an inter-organizational platform to support security management decisions. In *10th International Symposium on Human Aspects of Information Security & Assurance, HAISA 2016 ,Frankfurt, Germany, July 19-21, 2016, Proceedings.*, 2016. URL https://www.cscan.org/openaccess/?paperid=295.

[47] Julian Dax, Ana Ivan, Benedikt Ley, Sebastian Pape, Volkmar Pipek, Kai Rannenberg, Christopher Schmitz, and André Sekulla. IT security status of German energy providers. Technical report, Cornell University, arXiv, September 2017. URL https://arxiv.org/abs/1709.01254.

[48] Julian Dax, Benedikt Ley, Sebastian Pape, Volkmar Pipek, Kai Rannenberg, Christopher Schmitz, and André Sekulla. Stand zur IT-Sicherheit deutscher Stromnetzbetreiber : technischer Bericht, August 2017. URL https://dokumentix.ub.uni-siegen.de/opus/volltexte/2017/1185. Also available via https://arxiv.org/abs/1709.01254.

[49] Julian Dax, Daniel Hamburg, Sebastian Pape, Volkmar Pipek, Kai Rannenberg, Christopher Schmitz, André Sekulla, and Frank Terhaag. Sichere Informationsnetze bei kleinen und mittleren Energieversorgern (SIDATE). In S. Rudel and U. Lechner, editors, *State of the Art: IT-Sicherheit für Kritische Infrastrukturen*, chapter Sichere Informationsnetze bei kleinen und mittleren Energieversorgern (SIDATE), page 29. Universität der Bundeswehr, Neubiberg, 2018. URL https://www.itskritis.de/_uploads/user/IT-Sicherheit%20Kritische%20Infrastrukturen%E2%80%93screen.pdf.

[50] Julian Dax, Benedikt Ley, Sebastian Pape, Volkmar Pipek, Kai Rannenberg, Christopher Schmitz, and André Sekulla. Stand der IT-Sicherheit bei deutschen Stromnetzbetreibern. In S. Rudel and U. Lechner, editors, *State of the Art: IT-Sicherheit für Kritische Infrastrukturen*, chapter Stand der IT-Sicherheit bei deutschen Stromnetzbetreibern, pages 69–74. Universität der Bundeswehr, Neubiberg, 2018. URL https://www.itskritis.de/_uploads/user/IT-Sicherheit%20Kritische%20Infrastrukturen%E2%80%93screen.pdf.

[51] Julian Dax, Sebastian Pape, Volkmar Pipek, Kai Rannenberg, Christopher Schmitz, André Sekulla, and Frank Terhaag. Das SIDATE-Portal im Einsatz. In S. Rudel and U. Lechner, editors, *State of the Art: IT-Sicherheit für Kritische Infrastrukturen*, chapter Das SIDATE-Portal im Einsatz, pages 145–150. Universität der Bundeswehr, Neubiberg, 2018. URL https://www.itskritis.de/_uploads/user/IT-Sicherheit%20Kritische%20Infrastrukturen%E2%80%93screen.pdf.

[52] Tamara Denning, Adam Lerner, Adam Shostack, and Tadayoshi Kohno. Control-alt-hack: The design and evaluation of a card game for computer security awareness and education. In *Proceedings of the 2013 ACM SIGSAC Conference on Computer; Communications Security*, CCS '13, pages 915–928. ACM, 2013.

[53] Trajce Dimkov, André van Cleeff, Wolter Pieters, and Pieter Hartel. Two methodologies for physical penetration testing using social engineering. In *Proceedings of the 26th Annual Computer Security Applications Conference*, ACSAC '10, pages 399–408. ACM, 2010.

[54] Tamara Dinev and Paul Hart. An extended privacy calculus model for e-commerce transactions. *Information Systems Research*, 17(1):61–80, 2006.

[55] Roger Dingledine, Nick Mathewson, and Paul Syverson. Tor: The second-generation onion router. Technical report, Naval Research Lab Washington DC, 2004.

[56] Lynda Donovan and Pedagogical Lead. The use of serious games in the corporate sector. *A State of the Art Report. Learnovate Centre (December 2012)*, 2012.

[57] Nick Doty and Mohit Gupta. Privacy design patterns and anti-patterns patterns misapplied and unintended consequences. In *Trustbusters Workshop at the Symposium on Usable Privacy and Security*, 2013.

[58] Cynthia Dwork. Differential privacy: A survey of results. In *International conference on theory and applications of models of computation*, pages 1–19. Springer, 2008.

[59] Michael J Earl. The risks of outsourcing it. *Sloan management review*, 37:26–32, 1996.

[60] Peter Eckersley. How unique is your web browser? In *International Symposium on Privacy Enhancing Technologies Symposium*, volume 6205, pages 1–18. Springer, 2010.

[61] European Parliament and Council of The European Union. Regulation (EU) 2016/679 General Data Protection Regulation (GDPR), 2016. URL `http://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679&from=DE`.

[62] Shamal Faily and Ivan Flechais. Persona cases: a technique for grounding personas. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2267–2270, 2011.

[63] Martin Fishbein and Icek Ajzen. *Belief, Attitude, Intention and Behavior: An Introduction to Theory and Research*. Addison-Wesley, Reading, MA, 1975. doi: 10.2307/2065853.

[64] Alisa Frik and Alexia Gaudeul. The relation between privacy protection and risk attitudes, with a new experimental method to elicit the implicit monetary value of privacy. *CEGE Discussion Papers, Number*, 2016.

[65] Tal Garfinkel and Mendel Rosenblum. A virtual machine introspection based architecture for intrusion detection. In *In Proc. Network and Distributed Systems Security Symposium*, pages 191–206, 2003.

[66] Saurabh Kumar Garg, Steve Versteeg, and Rajkumar Buyya. SMICloud: A Framework for Comparing and Ranking Cloud Services. In *Proceedings of the 2011 Fourth IEEE International Conference on Utility and Cloud Computing*, UCC '11, pages 210–218, Washington, DC, USA, 2011. IEEE Computer Society. ISBN 978-0-7695-4592-9. doi: 10.1109/UCC.2011.36. URL `http://dx.doi.org/10.1109/UCC.2011.36`.

[67] Nina Gerber, Paul Gerber, and Melanie Volkamer. Explaining the privacy paradox: A systematic review of literature investigating privacy attitude and behavior. *Comput. Secur.*, 77:226–261, 2018. doi: 10.1016/j.cose.2018.04.002. URL `https://doi.org/10.1016/j.cose.2018.04.002`.

[68] Nina Gerber, Benjamin Reinheimer, and Melanie Volkamer. Investigating people's privacy risk perception. *Proc. Priv. Enhancing Technol.*, 2019(3):267–288, 2019. doi: 10.2478/popets-2019-0047. URL `https://doi.org/10.2478/popets-2019-0047`.

[69] Nina Gerber, Verena Zimmermann, and Melanie Volkamer. Why Johnny fails to protect his privacy. In *2019 IEEE European Symposium on Security and Privacy Workshops, EuroS&P Workshops 2019, Stockholm, Sweden, June 17-19, 2019*, pages 109–118. IEEE, 2019. ISBN 978-1-7281-3026-2. doi: 10.1109/EuroSPW.2019.00019. URL `https://doi.org/10.1109/EuroSPW.2019.00019`.

[70] Nirnay Ghosh, Soumya K. Ghosh, and Sajal K. Das. SelCSP: A Framework to Facilitate Selection of Cloud Service Providers. *IEEE Transactions on Cloud Computing*, 3(1):66–79, 2014. ISSN 2168-7161. doi: 10.1109/TCC.2014.2328578.

[71] Barney G. Glaser and Anselm L. Strauss. *The Discovery of Grounded Theory*. Aldine Pub., Chicago, 1967.

[72] Ludger Goeke, Alejandro Quintanar, Kristian Beckers, and Sebastian Pape. PROTECT - an easy configurable serious game to train employees against social engineering attacks. In *Computer Security - ESORICS 2019 International Workshops, IOSec, MSTEC, and FINSEC, Luxembourg City, Luxembourg, September 26-27, 2019, Revised Selected Papers*, volume 11981 of *Lecture Notes in Computer Science*, pages 156–171, Cham, 2019. Springer International Publishing. ISBN 978-3-030-42051-2. doi: 10.1007/978-3-030-42051-2_11. URL https://link.springer.com/chapter/10.1007/978-3-030-42051-2_11.

[73] Cornelia Graf, Peter Wolkerstorfer, Arjan Geven, and Manfred Tscheligi. A pattern collection for privacy enhancing technology. In *The 2nd Int. Conf. on Pervasive Patterns and Applications (PATTERNS 2010)*, pages 21–26, 2010.

[74] NIST Cloud Computing Security Working Group et al. NIST cloud computing security reference architecture. Technical report, National Institute of Standards and Technology, 2013.

[75] Radha Gulati. The threat of social engineering and your defense against it. *SANS Reading Room*, 2003.

[76] Sheikh Mahbub Habib, Sebastian Ries, Max Mühlhäuser, and Prabhu Varikkattu. Towards a trust management system for cloud computing marketplaces: using CAIQ as a trust information source. *Security and Communication Networks*, 7(11):2185–2200, 2014. ISSN 1939-0122. doi: 10.1002/sec.748. URL http://dx.doi.org/10.1002/sec.748.

[77] Thomas Haeberlen and Lionel Dupré. Cloud Computing - Benefits, Risks and Recommendations For Information Security. Technical report, ENISA, 2012.

[78] J. Hair, Christian M. Ringle, and Marko Sarstedt. PLS-SEM: Indeed a Silver Bullet. *Journal of Marketing Theory and Practice*, 19(2):139–152, 2011.

[79] David Harborth and Sebastian Pape. Examining technology use factors of privacy-enhancing technologies: The role of perceived anonymity and trust. In *24th Americas Conference on Information Systems, AMCIS 2018, New Orleans, LA, USA, August 16-18, 2018*. Association for Information Systems, 2018. URL https://aisel.aisnet.org/amcis2018/Security/Presentations/15.

[80] David Harborth and Sebastian Pape. JonDonym users' information privacy concerns. In *ICT Systems Security and Privacy Protection - 33rd IFIP TC 11 International Conference, SEC 2018, Held at the 24th IFIP World Computer Congress, WCC 2018, Poznan, Poland, September 18-20, 2018, Proceedings*, pages 170–184, 2018. doi: 10.1007/978-3-319-99828-2_13. URL https://doi.org/10.1007/978-3-319-99828-2_13.

[81] David Harborth and Sebastian Pape. German translation of the concerns for information privacy (CFIP) construct. Technical report, SSRN, January 2018. URL https://ssrn.com/abstract=3112207.

[82] David Harborth and Sebastian Pape. German translation of the unified theory of acceptance and use of technology 2 (UTAUT2) questionnaire. Technical report, SSRN, March 2018. URL https://ssrn.com/abstract=3147708.

[83] David Harborth and Sebastian Pape. How privacy concerns and trust and risk beliefs influence users' intentions to use privacy-enhancing technologies – the case of Tor. In *52nd Hawaii International Conference on System Sciences (HICSS) 2019*, pages 4851–4860, 01 2019. doi: 10125/59923. URL https://scholarspace.manoa.hawaii.edu/bitstream/10125/59923/1/0483.pdf.

[84] David Harborth and Sebastian Pape. How privacy concerns, trust and risk beliefs and privacy literacy influence users' intentions to use privacy-enhancing technologies - the case of Tor. *ACM SIGMIS Database: the DATABASE for Advances in Information Systems*, 51(1):51–69, January 2020. ISSN 0095-0033. doi: 10.1145/3380799.3380805. URL https://dl.acm.org/doi/abs/10.1145/3380799.3380805.

[85] David Harborth and Sebastian Pape. Dataset on actual users of the privacy-enhancing technology JonDonym, 2020. URL https://ieee-dataport.org/open-access/dataset-actual-users-privacy-enhancing-technology-jondonym.

[86] David Harborth and Sebastian Pape. Dataset on actual users of the privacy-enhancing technology Tor, 2020. URL https://ieee-dataport.org/open-access/dataset-actual-users-privacy-enhancing-technology-tor.

[87] David Harborth, Dominik Herrmann, Stefan Köpsell, Sebastian Pape, Christian Roth, Hannes Federrath, Dogan Kesdogan, and Kai Rannenberg. Integrating privacy-enhancing technologies into the internet infrastructure. Technical report, Cornell University, arXiv, November 2017. URL https://arxiv.org/abs/1709.01254. Also available via https://epub.uni-regensburg.de/36346/.

[88] David Harborth, Maren Braun, Akos Grosz, Sebastian Pape, and Kai Rannenberg. Anreize und Hemmnisse für die Implementierung von Privacy-Enhancing Technologies im Unternehmenskontext. In *Sicherheit 2018: Sicherheit, Schutz und Zuverlässigkeit, Beiträge der 9. Jahrestagung des Fachbereichs Sicherheit der Gesellschaft für Informatik e.V. (GI), 25.-27. April 2018, Konstanz*, pages 29–41, 2018. doi: 10.18420/sicherheit2018_02. URL https://doi.org/10.18420/sicherheit2018_02.

[89] David Harborth, Xinyuan Cai, and Sebastian Pape. Why do people pay for privacy? In *ICT Systems Security and Privacy Protection - 34th IFIP TC 11 International Conference, SEC 2019, Lisbon, Portugal, June 25-27, 2019, Proceedings*, pages 253–267, 06 2019. doi: 10.1007/978-3-030-22312-0_18. URL https://doi.org/10.1007/978-3-030-22312-0_18.

[90] David Harborth, Sebastian Pape, and Kai Rannenberg. Explaining the technology use behavior of privacy-enhancing technologies: The case of Tor and JonDonym. *Proceedings on Privacy Enhancing Technologies (PoPETs)*, 2020(2):111–128, 2020. doi: 10.2478/popets-2020-0020. URL https://content.sciendo.com/view/journals/popets/2020/2/article-p111.xml.

[91] Majid Hatamian and Jetzabel Serna-Olvera. Beacon alarming: Informed decision-making supporter and privacy risk analyser in smartphone applications. In *IEEE International Conference on Consumer Electronics, ICCE 2017, Las Vegas, NV, USA, January 8-10, 2017*, 2017.

[92] Majid Hatamian, Sebastian Pape, and Kai Rannenberg. ESARA: A framework for enterprise smartphone apps risk assessment. In *ICT Systems Security and Privacy Protection - 34th IFIP TC 11 International Conference, SEC 2019, Lisbon, Portugal, June 25-27, 2019, Proceedings*, pages 165–179, 06 2019. doi: 10.1007/978-3-030-22312-0_12. URL https://doi.org/10.1007/978-3-030-22312-0_12.

[93] Joseph M Hatfield. Virtuous human hacking: The ethics of social engineering in penetration-testing. *computers & security*, 83:354–366, 2019.

[94] Vera Hazilov and Sebastian Pape. Systematic scenario creation for serious security-awareness games. In Ioana Boureanu, Constantin Câtâlin Drâgan, Mark Manulis, Thanassis Giannetsos, Christoforos Dadoyan, Panagiotis Gouvas, Roger A. Hallman, Shujun Li, Victor Chang, Frank Pallas, Jörg Pohle, and Angela Sasse, editors, *Computer Security - ESORICS 2020 International Workshops, DETIPS, DeSECSys, MPS, and SPOSE, Guildford, UK, September 17-18, 2020, Revised Selected Papers*, volume 12580 of *LNCS*, pages 294–311, Cham, 09 2020. Springer International Publishing. doi: 10.1007/978-3-030-66504-3_18. URL https://link.springer.com/chapter/10.1007/978-3-030-66504-3_18.

[95] Jon Heales, Sophie Cockcroft, and Van-Hau Trieu. The influence of privacy, trust, and national culture on internet transactions. In Gabriele Meiselwitz, editor, *Social Computing and Social Media. Human Behavior*, pages 159–176, Cham, 2017. Springer International Publishing. ISBN 978-3-319-58559-8.

[96] Richard Henson and Bruce Hallas. SMEs, Information Risk Management, and ROI, 2009.

[97] Dominik Herrmann, Jens Lindemann, Ephraim Zimmer, and Hannes Federrath. Anonymity online for everyone: What is missing for zero-effort privacy on the internet? In *International Workshop on Open Problems in Network Security*, pages 82–94. Springer, 2015.

[98] Charles G Hill, Maren Haag, Alannah Oleson, Chris Mendez, Nicola Marsden, Anita Sarma, and Margaret Burnett. Gender-inclusiveness personas vs. stereotyping: Can we have it both ways? In *Proceedings of the 2017 chi conference on human factors in computing systems*, pages 6658–6671, 2017.

[99] Identify Theft Resource Center. Data breach reports, May 2020. URL https://www.idtheftcenter.org/wp-content/uploads/2020/06/May-2020-Data-Breach-Package.pdf.

[100] Information Commissioner's Office. Privacy regulators study finds Internet of Things shortfalls, 2016. URL https://ico.org.uk/about-the-ico/news-and-events/news-and-blogs/2016/09/privacy-regulators-study-finds-internet-of-things-shortfalls/.

[101] Information Systems Audit and Control Association (ISACA). *CobiT 5: A Business Framework for the Governance and Management of Enterprise IT*. ISACA, Rolling Meadows, 2012.

[102] International Organization for Standardization (ISO). ISO 23601:2009 safety identification – escape and evacuation plan signs. https://www.iso.org/standard/41685.html, 2009.

[103] International Organization for Standardization (ISO) and International Electrotechnical Commission (IEC). ISO/IEC 15504-5:2012, information technology – process assessment — part 5: An exemplar software life cycle process assessment model, 2012.

[104] International Organization for Standardization (ISO) and International Electrotechnical Commission (IEC). ISO/IEC 27001:2013 information technology – security techniques – information security management systems – requirements, October 2013.

[105] International Organization for Standardization (ISO) and International Electrotechnical Commission (IEC). ISO/IEC 27019:2017, information technology – security techniques – information security controls for the energy utility industry, October 2017.

[106] International Organization for Standardization (ISO) and International Electrotechnical Commission (IEC). ISO/IEC 27701:2019 Security techniques - Extension to ISO/IEC 27001 and ISO/IEC 27002 for privacy information management - Requirements and guidelines, 2019.

[107] Michaela Iorga, Larry Feldman, Robert Barton, Michael J Martin, Nedim S Goren, and Charif Mahmoudi. Fog computing conceptual model. Technical report, National Institute of Standards and Technology (NIST), 2018. NIST SP-500-325.

[108] Markus Jakobsson, Erik Stolterman, Susanne Wetzel, and Liu Yang. Love and authentication. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 197–200, 2008.

[109] Wayne Jansen and Timothy Grance. SP 800-144. guidelines on security and privacy in public cloud computing. Technical report, NIST, 2011.

[110] JonDos Gmbh. Official Homepage of JonDonym. https://www.anonym-surfen.de, 2018.

[111] Mike Just. Designing and evaluating challenge-question systems. *IEEE Security & Privacy*, 2(5): 32–39, 2004.

[112] Dennis-Kenji Kipker. The EU NIS directive compared to the IT security act - Germany is well positioned for the new European cybersecurity space. *ZD-Aktuell*, 6(20):05363, 2016.

[113] Dennis-Kenji Kipker, Sebastian Pape, Stefanie Wojak, and Kristian Beckers. Juristische Bewertung eines Social-Engineering-Abwehr Trainings. In S. Rudel and U. Lechner, editors, *State of the Art: IT-Sicherheit für Kritische Infrastrukturen*, chapter Stand der IT-Sicherheit bei deutschen Stromnetzbetreibern, pages 112–115. Universität der Bundeswehr, Neubiberg, 2018. URL https://www.itskritis.de/_uploads/user/IT-Sicherheit%20Kritische%20Infrastrukturen%E2%80%93screen.pdf.

[114] Michael S. Kirkpatrick, Gabriel Ghinita, and Elisa Bertino. Resilient authenticated execution of critical applications in untrusted environments. *IEEE Transactions on Dependable and Secure Computing*, 9 (4):597–609, 2012. ISSN 1545-5971. doi: 10.1109/TDSC.2012.25.

[115] Paul Kocher. Timing attacks on implementations of Diffie-Hellman, RSA, DSS, and other systems. In *Annual International Cryptology Conference*, pages 104–113. Springer, 1996.

[116] Paul Kocher, Jann Horn, Anders Fogh, Daniel Genkin, Daniel Gruss, Werner Haas, Mike Hamburg, Moritz Lipp, Stefan Mangard, Thomas Prescher, et al. Spectre attacks: Exploiting speculative execution. In *2019 IEEE Symposium on Security and Privacy (SP)*, pages 1–19. IEEE, 2019.

[117] Hristo Koshutanski, Marinos Tsantekidis, Ernesto Damiani, Fulvio Frati, Stelvio Cimato, Elvinia Riccobene, George Hatzivasilis, Konstantinos Fysarakis, George Spanoudakis, Oleg Blinder, Michael Vinov, Torsten Hildebrandt, Dirk Wortmann, Vina Rompoti, George Bravos, Vassilis Chatzigiannakis, Kristian Beckers, Sebastian Pape, Martin Kunc, and Pavel Bašta. THREAT-ARREST platform's initial reference architecture. Technical report, Threat-Arrest, 2019. Deliverable 1.3.

[118] Katharina Krombholz, Heidelinde Hobel, Markus Huber, and Edgar Weippl. Social Engineering Attacks on the Knowledge Worker Categories and Subject Descriptors. *Security Information and Networks*, pages 28–35, 2013. doi: 10.1145/2523514.2523596.

[119] Hennie A. Kruger and Wayne D. Kearney. A prototype for assessing information security awareness. *Computers & security*, 25(4):289–296, 2006.

[120] Ronald L. Krutz and Russell Dean Vines. *Cloud security: a comprehensive guide to secure cloud computing*. Wiley, 2010. ISBN 9780470589878.

[121] Jörn Kuhn and Alexander Willemsen. Arbeitsrechtliche Aspekte von Social Engineering Audits. *DER BETRIEB*, 02:111–117, 2016. URL `https://www.wiso-net.de/document/MCDB__DBDBDB1167400`.

[122] Pierre Laperdrix, Walter Rudametkin, and Benoit Baudry. Beauty and the beast: Diverting modern web browsers to build unique browser fingerprints. In *IEEE S&P*, 2016.

[123] Linked.In. Quick android review kit, 2020. URL `https://github.com/linkedin/qark`.

[124] Federico Maggi, Andrea Valdi, and Stefano Zanero. Andrototal: A flexible, scalable toolbox and service for testing mobile malware detectors. In *Proc. of the 3rd ACM Workshop on Security and Privacy in Smartphones and Mobile Devices*, pages 49–54, 2013. ISBN 978-1-4503-2491-5.

[125] Naresh K. Malhotra, Sung S. Kim, and James Agarwal. Internet users' information privacy concerns (IUIPC): The construct, the scale, and a causal model. *Information Systems Research*, 15(4):336–355, 2004.

[126] Akshaya Mani, T. Wilson-Brown, Rob Jansen, Aaron Johnson, and Micah Sherr. Understanding Tor Usage with Privacy-Preserving Measurement. In *2018 Internet Measurement Conference (IMC'18)*, pages 1–13, 2018. ISBN 9781450356190. doi: 10.1145/3278532.3278549.

[127] Karola Marky, Andreas Gutmann, Philipp Rack, and Melanie Volkamer. Privacy friendly apps - making developers aware of privacy violations. In David Aspinall, Lorenzo Cavallaro, Mohamed Nassim Seghir, and Melanie Volkamer, editors, *Proceedings of the 1st International Workshop on Innovations in Mobile Privacy and Security, IMPS 2016, co-located with the International Symposium on Engineering Secure Software and Systems (ESSoS 2016), London, UK, April 6, 2016*, volume 1575 of *CEUR Workshop Proceedings*, pages 46–48. CEUR-WS.org, 2016. URL `http://ceur-ws.org/Vol-1575/paper_4.pdf`.

[128] Philipp K. Masur, Doris Teutsch, and Sabine Trepte. Entwicklung und Validierung der Online-Privatheitskompetenzskala (OPLIS). *Diagnostica*, 2017.

[129] Christoph P. Mayer. Security and Privacy Challenges in the Internet of Things. In *Electronic Communications of the EASST*, volume 17. European Association of Software Science and Technology, 2009.

[130] Nikola Milosevic. Introduction to Social Engineering, 2013. URL http://inspiratron.org/introduction-to-social-engineering/.

[131] Liz Mineo. On internet privacy, be very afraid (Interview with Bruce Schneier). https://news.harvard.edu/gazette/story/2017/08/when-it-comes-to-internet-privacy\-be-very-afraid-analyst-suggests/, 08 2017.

[132] Kevin D. Mitnick and William L. Simon. *The art of deception: Controlling the human element of security*. John Wiley & Sons, 2003.

[133] Antonio Montieri, Domenico Ciuonzo, Giuseppe Aceto, and Antonio Pescapé. Anonymity services Tor, I2P, JonDonym: Classifying in the dark. In *Teletraffic Congress (ITC 29), 2017 29th International*, volume 1, pages 81–89. IEEE, 2017.

[134] Philipp Morgner, Christoph Mai, Nicole Koschate-Fischer, Felix C. Freiling, and Zinaida Benenson. Security update labels: Establishing economic incentives for security patching of IoT consumer products. In *2020 IEEE Symposium on Security and Privacy, SP 2020, San Francisco, CA, USA, May 18-21, 2020*, pages 429–446, 2020. doi: 10.1109/SP40000.2020.00021. URL https://doi.org/10.1109/SP40000.2020.00021.

[135] Joachim Müller, Alexander Sänn, MdB Marian Wendt, RA Annett Albrecht, and Peter Langendörfer. Informationssicherheits-Management-Systeme (ISMS) bei Energieversorgern 2018. Technical report, Betriebswirtschaftliches Forschungszentrum für Fragen der mittelständischen Wirtschaft e. V. an der Universität Bayreuth and Seven Principles AG, 2018.

[136] Pardis Emami Naeini, Sruti Bhagavatula, Hana Habib, Martin Degeling, Lujo Bauer, Lorrie Cranor, and Norman Sadeh. Privacy expectations and preferences in an IoT world. In *Symposium on Usable Privacy and Security (SOUPS)*, 2017.

[137] National Electric Sector Cybersecurity Organization Resource (NESCOR). Electric sector failure scenarios and impact analyses. Technical report, National Electric Sector Cybersecurity Organization Resource (NESCOR), 2013.

[138] National Institute for Standards and Technology (NIST). NIST special publication 800-53 – security and privacy controls for federal information systems and organizations. http://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-53r4.pdf, April 2013.

[139] Donald A Norman. The research-practice gap: The need for translational developers. *Interactions*, 17 (4):9–12, 2010.

[140] nusenu. How malicious Tor relays are exploiting users in 2020 (part i). Medium, Aug 2020. URL https://medium.com/@nusenu/how-malicious-tor-relays-are-exploiting-users-in-2020-part-i-1097575c0cac.

[141] NVISO Labs. Nviso. apkscan., 2019. URL https://apkscan.nviso.be/. Service retired in October 2019.

[142] Osterlab. Mobile application security scanner, 2020. URL https://www.ostorlab.co/.

[143] Miaria Papadaki, Steven Furnell, and Ronald C. Dodge. Social engineering: Exploiting the weakest links. *European Network & Information Security Agency (ENISA), Heraklion, Crete*, 2008.

[144] Sebastian Pape. Sicherheitsmodelle für das Ajtai-Dwork-Kryptosystem, Januar 2004.

[145] Sebastian Pape. *Sicherheitsmodelle für das Ajtai-Dwork-Kryptosystem: Untersuchungen eines Kryptosystems mit Worst-Case / Average-Case Äquivalenz zum unique Shortest Vector Problem*. Vdm Verlag Dr. Müller, Oktober 2008. ISBN 978-3639043143.

[146] Sebastian Pape. Templateless biometric-enforced non-transferability of anonymous credentials (extended abstract). Technical report, Weimar, July 2008. URL https://nbn-resolving.de/urn/resolver.pl?urn=urn:nbn:de:bvb:473-opus-1333.

[147] Sebastian Pape. Embedding biometric information into anonymous credentials. Technical Report 68, February 2008.

[148] Sebastian Pape. A survey on non-transferable anonymous credentials. In Vashek Matyáš, Simone Fischer-Hübner, Daniel Cvrček, and Petr Švenda, editors, *The Future of Identity in the Information Society*, volume 298 of *IFIP Advances in Information and Communication Technology*, pages 107–118. Springer Boston, Brno, Czech Republic, July 2009. doi: 10.1007/978-3-642-03315-5_8.

[149] Sebastian Pape. Some observations on reusing one-time pads within dice codings (abstract). Technical report, March 2009.

[150] Sebastian Pape. The challenge of authentication in insecure environments, 2013. (defended, September 2nd, 2013).

[151] Sebastian Pape. *Authentication in Insecure Environments – Using Visual Cryptography and Non-Transferable Credentials in Practise*. Research. Springer Vieweg, 2014. doi: 10.1007/978-3-658-07116-5. URL https://www.springer.com/springer+vieweg/it+%26+informatik/datenbanken/book/978-3-658-07115-8. eBook ISBN 978-3-658-07116-5, Softcover ISBN 978-3-658-07115-8.

[152] Sebastian Pape. Sample or random security - A security model for segment-based visual cryptography. In *Financial Cryptography and Data Security - 18th International Conference, FC 2014, Christ Church, Barbados, March 3-7, 2014, Revised Selected Papers*, pages 291–303, 2014. doi: 10.1007/978-3-662-45472-5_19. URL https://link.springer.com/chapter/10.1007%2F978-3-662-45472-5_19.

[153] Sebastian Pape and Dennis-Kenji Kipker. Case study: Checking a serious security-awareness game for its legal adequacy. unpublished manuscript, 2020.

[154] Sebastian Pape and Kai Rannenberg. Applying privacy patterns to the internet of things' (IoT) architecture. *Mobile Networks and Applications (MONET) – The Journal of SPECIAL ISSUES on Mobility of Systems, Users, Data and Computing*, 24(3):925–933, 06 2019. doi: 10.1007/s11036-018-1148-2. URL https://doi.org/10.1007/s11036-018-1148-2.

[155] Sebastian Pape and Jelena Stankovic. An insight into decisive factors in cloud provider selection with a focus on security. In *Computer Security - ESORICS 2019 International Workshops, CyberICPS, SECPRE, SPOSE, ADIoT, Luxembourg City, Luxembourg, September 26-27, 2019, Revised Selected Papers*, volume 11980 of *Lecture Notes in Computer Science*, pages 287–306, Cham, 09 2019. Springer International Publishing. ISBN 978-3-030-42048-2. doi: 10.1007/978-3-030-42048-2_19. URL https://link.springer.com/chapter/10.1007/978-3-030-42048-2_19.

[156] Sebastian Pape, Volkmar Pipek, Kai Rannenberg, Christopher Schmitz, André Sekulla, and Frank Terhaag. Stand zur IT-Sicherheit deutscher Stromnetzbetreiber : technischer Bericht. Technical report, Universität Siegen, December 2018. URL https://dokumentix.ub.uni-siegen.de/opus/volltexte/2018/1394/.

[157] Sebastian Pape, Daniel Tasche, Iulia Bastys, Akos Grosz, Joerg Laessig, and Kai Rannenberg. Towards an architecture for pseudonymous e-commerce – applying privacy by design to online shopping. In *Sicherheit 2018: Sicherheit, Schutz und Zuverlässigkeit, Beiträge der 9. Jahrestagung des Fachbereichs Sicherheit der Gesellschaft für Informatik e.V. (GI), 25.-27. April 2018, Konstanz*, pages 17–28, 2018. doi: 10.18420/sicherheit2018_01. URL https://doi.org/10.18420/sicherheit2018_01.

[158] Sebastian Pape, Ludger Goeke, Alejandro Quintanar, and Kristian Beckers. Conceptualization of a cybersecurity awareness quiz. In *Computer Security - ESORICS 2020 International Workshops MSTEC*, volume 12512 of *LNCS*, pages 61–76, Cham, 09 2020. Springer International Publishing. doi: 10.1007/978-3-030-62433-0_4. URL https://link.springer.com/chapter/10.1007%2F978-3-030-62433-0_4.

[159] Sebastian Pape, Ana Ivan, David Harborth, Toru Nakamura, Shinsaku Kiyomoto, Haruo Takasaki, and Kai Rannenberg. Re-evaluating internet users' information privacy concerns: The case in japan. *AIS Transactions on Replication Research*, 6(18):1–18, 10 2020. doi: 10.17705/1atrr.00061. URL https://aisel.aisnet.org/trr/vol6/iss1/18/.

[160] Sebastian Pape, Ana Ivan, David Harborth, Toru Nakamura, Shinsaku Kiyomoto, Haruo Takasaki, and Kai Rannenberg. Open materials discourse: Re-evaluating internet users' information privacy concerns: The case in japan. *AIS Transactions on Replication Research*, 6(22):1–7, 10 2020. doi: 10.17705/1atrr.00065. URL https://aisel.aisnet.org/trr/vol6/iss1/22.

[161] Sebastian Pape, Federica Paci, Jan Juerjens, and Fabio Massacci. Selecting a secure cloud provider: An empirical study and multi criteria approach. *Information*, 11(5), 05 2020. doi: 10.3390/info11050261. URL https://www.mdpi.com/2078-2489/11/5/261. Section Information Applications, Special Issue Cloud Security Risk Management.

[162] Sebastian Pape, Christopher Schmitz, Dennis-Kenji Kipker, and Andre Sekula. On the use of information security management systems by german energy providers. Accepted for presentation at the Fourteenth IFIP Working Group 11.10 International Conference on Critical Infrastructure Protection, 03 2020.

[163] Yong Jin Park. Digital literacy and privacy behavior online. *Communication Research*, 40(2):215–236, 2013.

[164] Ioannis Patiniotakis, Stamatia Rizou, Yiannis Verginadis, and Gregoris Mentzas. Managing imprecise criteria in cloud service ranking with a fuzzy multi-criteria decision making method. In Kung-Kiu Lau, Winfried Lamersdorf, and Ernesto Pimentel, editors, *Service-Oriented and Cloud Computing*, volume 8135 of *Lecture Notes in Computer Science*, pages 34–48. Springer Berlin Heidelberg, 2013. ISBN 978-3-642-40650-8. doi: 10.1007/978-3-642-40651-5_4. URL http://dx.doi.org/10.1007/978-3-642-40651-5_4.

[165] Niklas Paul, Welderufael B. Tesfay, Dennis-Kenji Kipker, Mattea Stelter, and Sebastian Pape. Assessing privacy policies of internet of things services. In *ICT Systems Security and Privacy Protection - 33rd IFIP TC 11 International Conference, SEC 2018, Held at the 24th IFIP World Computer Congress, WCC 2018, Poznan, Poland, September 18-20, 2018, Proceedings*, pages 156–169, 2018. doi: 10.1007/978-3-319-99828-2_12. URL https://doi.org/10.1007/978-3-319-99828-2_12.

[166] Thomas R. Peltier. Social engineering: Concepts and solutions. *Information Security Journal*, 15(5): 13, 2006.

[167] Mike Perry, Erinn Clark, and Steven Murdoch. The design and implementation of the tor browser. Technical report, The Tor Project, 2013. https://www.torproject.org/projects/torbrowser/design/.

[168] Andreas Pfitzmann and Marit Hansen. A terminology for talking about privacy by data minimization: Anonymity, unlinkability, undetectability, unobservability, pseudonymity, and identity management, 2010.

[169] Clint Pollock. Mobile app top 10, 2016. URL https://owasp.org/www-pdf-archive/MobileTopTen.pdf.

[170] Sören Preibusch, Thomas Peetz, Gunes Acar, and Bettina Berendt. Shopping for privacy: Purchase details leaked to paypal. *Electronic Commerce Research and Applications*, 2016.

[171] Pearl Pu, Li Chen, and Rong Hu. A user-centric evaluation framework for recommender systems. In *Proceedings of the fifth ACM conference on Recommender systems*, pages 157–164. ACM, 2011.

[172] James Brian Quinn and Frederick G Hilmer. Strategic outsourcing. *MIT Sloan Management Review*, 35(4):43, 1994.

[173] Quixxi Corp. Quixxi integrated app management system, 2020. URL `https://quixxisecurity.com/`.

[174] Frederic Raber and Antonio Krueger. Towards understanding the influence of personality on mobile app permission settings. In *IFIP Conference on Human-Computer Interaction*, pages 62–82. Springer, 2017.

[175] Ariel Rabkin. Personal knowledge questions for fallback authentication: Security questions in the era of facebook. In *Proceedings of the 4th symposium on Usable privacy and security*, pages 13–23, 2008.

[176] Rajasree K. Rajamma, Audhesh K. Paswan, and Muhammad M. Hossain. Why do shoppers abandon shopping cart? Perceived waiting time, risk, and transaction inconvenience. *Journal of Product & Brand Management*, 2009.

[177] Christian M. Ringle, S. Wende, and Jan Michael Becker. SmartPLS 3. `www.smartpls.com`, 2015.

[178] Thomas Ristenpart, Eran Tromer, Hovav Shacham, and Stefan Savage. Hey, You, Get Off of My Cloud: Exploring Information Leakage in Third-Party Compute Clouds. In *CCS '09: Proceedings of the 16th ACM conference on Computer and Communications Security*, pages 199–212, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-894-0. doi: http://doi.acm.org/10.1145/1653662.1653687.

[179] Thomas L. Saaty. *Theory and applications of the analytic network process: decision making with benefits, opportunities, costs, and risks*. RWS publications, 2005.

[180] Thomas L. Saaty. Decision making with the analytic hierarchy process. *International journal of services sciences*, 1(1):83–98, 2008.

[181] Saad Saleh, Junaid Qadir, and Muhammad U. Ilyas. Shedding light on the dark corners of the internet: A survey of tor research. *Journal of Network and Computer Applications*, 114:1 – 28, 2018. ISSN 1084-8045. doi: https://doi.org/10.1016/j.jnca.2018.04.002. URL `http://www.sciencedirect.com/science/article/pii/S1084804518301280`.

[182] Tarik Saleh. Covidlock update: Deeper analysis of coronavirus android ransomware. `https://www.domaintools.com/resources/blog/covidlock-update-coronavirus-ransomware`, 2020.

[183] Peter Schaab, Kristian Beckers, and Sebastian Pape. A systematic gap analysis of social engineering defence mechanisms considering social psychology. In *10th International Symposium on Human Aspects of Information Security & Assurance, HAISA 2016, Frankfurt, Germany, July 19-21, 2016, Proceedings.*, 2016. URL `https://www.cscan.org/openaccess/?paperid=301`.

[184] Peter Schaab, Kristian Beckers, and Sebastian Pape. Social engineering defence mechanisms and counteracting training strategies. *Information and Computer Security*, 25(2):206–222, 2017. doi: 10.1108/ICS-04-2017-0022. URL `https://doi.org/10.1108/ICS-04-2017-0022`.

[185] Stuart Schechter, AJ Bernheim Brush, and Serge Egelman. It's no secret. measuring the security and reliability of authentication via "secret" questions. In *2009 30th IEEE Symposium on Security and Privacy*, pages 375–390. IEEE, 2009.

[186] Michael Schmid and Sebastian Pape. Aggregating corporate information security maturity levels of different assets. In *Privacy and Identity Management. Data for Better Living: AI and Privacy - 14th IFIP WG 9.2, 9.6/11.7, 11.6/SIG 9.2.2 International Summer School, Windisch, Switzerland, August 19-23, 2019, Revised Selected Papers*, number 576 in IFIP Advances in Information and Communication Technology, pages 376–392. Springer Boston, 2019. doi: 10.1007/978-3-030-42504-3_24. URL `https://link.springer.com/chapter/10.1007/978-3-030-42504-3_24`.

[187] Michael Schmid and Sebastian Pape. A structured comparison of the corporate information security. In *ICT Systems Security and Privacy Protection - 34th IFIP TC 11 International Conference, SEC 2019, Lisbon, Portugal, June 25-27, 2019, Proceedings*, pages 223–237, 06 2019. doi: 10.1007/978-3-030-22312-0_16. URL https://doi.org/10.1007/978-3-030-22312-0_16.

[188] Christopher Schmitz and Sebastian Pape. LiSRA: Lightweight security risk assessment for decision support in information security. *Computers & Security*, 90, 2020. doi: 10.1016/j.cose.2019.101656. URL https://www.sciencedirect.com/science/article/pii/S0167404819301993.

[189] Christopher Schmitz, Andre Sekula, Sebastian Pape, Volkmar Pipek, and Kai Rannenberg. Easing the burden of security self-assessments. In *12th International Symposium on Human Aspects of Information Security & Assurance, HAISA 2018, Dundee, Scotland, August 29-31, 2018, Proceedings.*, 2018.

[190] Christopher Schmitz, André Sekulla, and Sebastian Pape. Asset-centric analysis and visualisation of attack trees. In *Graphical Models for Security - 7th International Workshop, GraMSec@CSF 2020, Boston, MA, USA, Virtual Conference, June 22, 2020, Revised Selected Papers*, volume 12419 of *LNCS*, pages 45–64. Springer, 11 2020. doi: 10.1007/978-3-030-62230-5_3. URL https://link.springer.com/chapter/10.1007%2F978-3-030-62230-5_3.

[191] Donald J. Schuirmann. A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. *Journal of pharmacokinetics and biopharmaceutics*, 15(6):657–680, 1987.

[192] André Sekulla, Christopher Schmitz, Sebastian Pape, and Volkmar Pipek. Demonstrator zur Beschreibung und Visualisierung einer kritischen Infrastruktur. In *Human Practice. Digital Ecologies. Our Future. 14. Internationale Tagung Wirtschaftsinformatik (WI 2019), February 24-27, 2019, Siegen, Germany*, page 1978, 02 2019. URL https://wi2019.de/wp-content/uploads/Tagungsband_WI2019_reduziert.pdf.

[193] Anirban Sengupta. Modeling dependencies of ISO/IEC 27002:2013 security controls. In Jemal H. Abawajy, Sougata Mukherjea, Sabu M. Thampi, and Antonio Ruiz-Martínez, editors, *Security in Computing and Communications*, pages 354–367, Cham, 2015. Springer International Publishing. ISBN 978-3-319-22915-7.

[194] Andrea Shalal, Thorsten Severin, and Christoph Steitz. German prosecutors probing hack of energy firm last year. Reuters Technology News, 2018. URL https://www.reuters.com/article/us-germany-energy-cyber/german-prosecutors-probing-hack-of-energy-firm-last-year-idUSKCN1IH1QD.

[195] Steve Sheng, Levi Broderick, Colleen Alison Koranda, and Jeremy J Hyland. Why Johnny still can't encrypt: evaluating the usability of email encryption software. In *Symposium On Usable Privacy and Security*, pages 3–4. ACM, 2006.

[196] Christopher Shields. Aristotle's psychology. *Stanford Encyclopedia of Philosophy*, 2000.

[197] Adam Shostack. Elevation of privilege: Drawing developers into threat modeling. Technical report, Microsoft, Redmond, U.S., 2012. http://download.microsoft.com/download/F/A/E/FAE1434F-6D22-4581-9804-8B60C04354E4/EoP_Whitepaper.pdf.

[198] Adam Shostack. *Threat Modeling: Designing for Security*. John Wiley & Sons Inc., 1st edition, 2014.

[199] Amit Kumar Sikder, Giuseppe Petracca, Hidayet Aksu, Trent Jaeger, and A. Selcuk Uluagac. A survey on sensor-based threats to internet-of-things (iot) devices and applications. *CoRR*, abs/1802.02041, 2018. URL http://arxiv.org/abs/1802.02041.

[200] Frank Stajano and Paul Wilson. Understanding scam victims: Seven principles for systems security. *Commun. ACM*, 54(3):70–75, March 2011. ISSN 0001-0782. doi: 10.1145/1897852.1897872. URL http://doi.acm.org/10.1145/1897852.1897872.

[201] Christoph Steitz and Eric Auchard. German nuclear plant infected with computer viruses, operator says. Reuters Technology News, 2016. URL https://www.reuters.com/article/us-nuclearpower-cyber-germany/german-nuclear-plant-infected-with-computer-viruses-operator-says-idUSKCN0XN2OS.

[202] Kathy A. Stewart and Albert H. Segars. An Empirical Examination of the Concern for Information Privacy Instrument. *Information Systems Research*, 13(1):36–49, 2002.

[203] Smitha Sundareswaran, Anna Squicciarini, and Dan Lin. A brokerage-based approach for cloud service selection. In *Cloud Computing (CLOUD), 2012 IEEE 5th International Conference on*, pages 558–565. IEEE, June 2012. doi: 10.1109/CLOUD.2012.119.

[204] Latanya Sweeney. k-anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10(05):557–570, 2002.

[205] Irfan Syamsuddin and Junseok Hwang. The application of AHP to evaluate information security policy decision making. *International Journal of Simulation: Systems, Science and Technology*, 10(4): 46–50, 2009. ISSN 14738031.

[206] The Tor Project. https://www.torproject.org, 2018.

[207] Jean Tirole. Cognition and incomplete contracts. *American Economic Review*, 99(1):265–94, 2009. doi: 10.1257/aer.99.1.265. URL http://www.aeaweb.org/articles.php?doi=10.1257/aer.99.1.265.

[208] Sabine Trepte, Doris Teutsch, Philipp K. Masur, Carolin Eicher, Mona Fischer, Alisa Hennhöfer, and Fabienne Lind. Do people know about privacy and data protection strategies? towards the online privacy literacy scale (oplis). In *Reforming European data protection law*, pages 333–365. Springer, 2015.

[209] Markus Tschersich, Shinsaku Kiyomoto, Sebastian Pape, Toru Nakamura, Gökhan Bal, Haruo Takasaki, and Kai Rannenberg. On gender specific perception of data sharing in japan. In *ICT Systems Security and Privacy Protection - 31st IFIP TC 11 International Conference, SEC 2016, Ghent, Belgium, May 30 - June 1, 2016, Proceedings*, pages 150–160, 2016. doi: 10.1007/978-3-319-33630-5_11. URL https://dx.doi.org/10.1007/978-3-319-33630-5_11.

[210] Melanie Volkamer, Martina Angela Sasse, and Franziska Boehm. Analysing simulated phishing campaigns for staff. In *Computer Security - ESORICS 2020 International Workshops, 2nd Workshop on Security, Privacy, Organizations, and Systems Engineering (SPOSE), to appear*, 09 2020.

[211] Gavin Watson, Andrew Mason, and Richard Ackroyd. *Social Engineering Penetration Testing*. Elsevier, 2014. ISBN 9780124201248. doi: 10.1016/B978-0-12-420124-8.00012-0. URL http://www.sciencedirect.com/science/article/pii/B9780124201248000120.

[212] Alma Whitten and J Doug Tygar. Why Johnny can't encrypt: A usability evaluation of pgp 5.0. In *USENIX Security Symposium*, volume 348, pages 169–184, 1999.

[213] Dwayne Whitten and Dorothy Leidner. Bringing it back: An analysis of the decision to backsource or switch vendors. *Decision Sciences*, 37(4):605–621, 2006.

[214] Laurie Williams, Michael Gegick, and Andrew Meneely. Protection poker: Structuring software security risk assessment and knowledge transfer. In *Proceedings of International Symposium on Engineering Secure Software and Systems*, pages 122–134. Springer, 2009.

[215] Laurie Williams, Andrew Meneely, and Grant Shipley. Protection poker: The new software security "game". *IEEE Security & Privacy*, 8(3):14–20, May 2010. ISSN 1540-7993.

[216] Erik Wittern, Jörn Kuhlenkamp, and Michael Menzel. Cloud service selection based on variability modeling. In Chengfei Liu, Heiko Ludwig, Farouk Toumani, and Qi Yu, editors, *Service-Oriented Computing*, volume 7636 of *Lecture Notes in Computer Science*, pages 127–141. Springer Berlin Heidelberg, 2012.

[217] Kim Wuyts and Wouter Joosen. LINDDUN privacy threat modeling: a tutorial, 2015.

[218] Alison Jing Xu and Robert S. Wyer Jr. The role of bolstering and counterarguing mind-sets in persuasion. *Journal of Consumer Research*, 38(5):920–932, 2012.

[219] Shanhe Yi, Zhengrui Qin, and Qun Li. Security and privacy issues of fog computing: A survey. In *International conference on wireless algorithms, systems, and applications*, pages 685–695. Springer, 2015.

[220] Mark Zimmer and Alicia Helle. Tests mit Tücke - Arbeitsrechtliche Anforderungen an Social Engineering Tests. *Betriebs-Berater*, 21/2016:1269, 2016.

All URLs have been last accessed on September 25th, 2020.

Appendix A

Social Engineering

A.1 A Serious Game for Eliciting Social Engineering Security Requirements

Kristian Beckers and Sebastian Pape. A serious game for eliciting social engineering security requirements. In *Proceedings of the 24th IEEE International Conference on Requirements Engineering*, RE '16. IEEE Computer Society, 2016. doi: 10.1109/RE.2016.39

A Serious Game for Eliciting Social Engineering Security Requirements

Kristian Beckers
Technische Universität München,
Germany
Email: beckersk@in.tum.de

Sebastian Pape
Goethe-University Frankfurt
Germany
Email: sebastian.pape@m-chair.de

***Index Terms*—security requirements elicitation, requirements prioritisation, threat analysis, gamification**

***Abstract*—Social engineering is the acquisition of information about computer systems by methods that deeply include non-technical means. While technical security of most critical systems is high, the systems remain vulnerable to attacks from social engineers. Social engineering is a technique that: (i) does not require any (advanced) technical tools, (ii) can be used by anyone, (iii) is cheap.**

Traditional security requirements elicitation approaches often focus on vulnerabilities in network or software systems. Few approaches even consider the exploitation of humans via social engineering and none of them elicits personal behaviours of individual employees. While the amount of social engineering attacks and the damage they cause rise every year, the security awareness of these attacks and their consideration during requirements elicitation remains negligible.

We propose to use a card game to elicit these requirements, which all employees of a company can play to understand the threat and document security requirements. The game considers the individual context of a company and presents underlying principles of human behaviour that social engineers exploit, as well as concrete attack patterns. We evaluated our approach with several groups of researchers, IT administrators, and professionals from industry.

I. Introduction

"The biggest threat to the security of a company is not a computer virus, an unpatched hole in a key program or a badly installed firewall. In fact, the biggest threat could be you [...] What I found personally to be true was that it's easier to manipulate people rather than technology [...] Most of the time organizations overlook that human element."[1] These words from Kevin Mitnick spoken in a BBC interview were made over a decade ago and are still of utmost importance today. A Dimensional Research study[2] with 853 IT professionals from United States, United Kingdom, Canada, Australia, New Zealand, and Germany about social engineering in 2011 confirmed Mitnick's statement. It revealed that 48% of large companies and 32% of small companies fell victim to 25 or more social engineering attacks in the past two years. The average cost per incident was over $25 000. 30% of large companies even cited a per incident cost of over $100 000. The SANS institute released a white paper[3] with even more severe numbers about social engineering. It states that cyber attacks cost U.S. companies $266 million every year and that 80% of all attacks are caused by authorized users. These users are either disgruntled employees or non-employees that have established trust within a company.

Eliciting security requirements for human threats is essential to consider the right defense mechanisms for concerns of socio-technical systems (STS). This elicitation is difficult for security engineers, because these are trained to focus mainly on other aspects of STS such as business processes, software applications, and hardware components. Additionally, external security engineers would have to gather relevant domain knowledge to understand the company, e.g. learn about processes, policies, employees' capabilities and attitudes. A common theme in security requirements engineering is modeling aspects of STS. For example, Lamsweerde [1] investigates security requirements for software, Mouratidis [2] and Liu [3] analyze organizational security issues, and Herrmann [4] focuses on business processes. The work of Li [5] considers all aspects of STS in one holistic model. These approaches have in common that they often assume the security requirements are known by the stakeholders and have only to be made explicit via modeling. This leads to a gap in the security analysis if the stakeholders are not aware of social engineering threats. Some approaches use patterns to identify threats [6], [7], which is generally a good idea, but for social engineering difficult, since the personality traits of individual persons such as *writes passwords on post-it notes* have to be known and described in a model. That is currently not done in security requirements engineering.

Several approaches focus on the elicitation of security requirements in different ways. Houmb [8] uses the Common Criteria as a basis for identifying security concerns in software documentation, Herrmann [9] relies on business risks for eliciting security requirements. These approaches build on existing software and business documentation as a source for security requirements, which does not focus on the behavior of humans in a company that might be exploited by a software engineer. Several works propose to use brainstorming as a

[1] news.bbc.co.uk/2/hi/technology/2320121.stm

[2] http://docplayer.net/11092603-The-risk-of-social-engineering-on-information-security.html

[3] http://www.sans.org/reading-room/whitepapers/engineering/threat-social-engineering-defense-1232

source for security requirements, e.g., Ionita [10]. These may result in social engineering security requirements, but again only if the stakeholders come up with the idea of social engineering, which requires them to know about it beforehand.

Recently, serious games have built reputation for getting employees of companies involved in security activities in an enjoyable and sustainable way. While still preserving a playful character, serious games are designed for a primary purpose other than pure entertainment, e.g. education, awareness training, social change. Williams et al. [11], [12] introduced the protection poker game to prioritize risks in software engineering projects. Shostack [13], [14] from Microsoft presented his Elevation of Privileges (EoP) card game to practice threat analysis with software engineers. We believe a serious game is relevant for social engineering, as well. Furthermore, games are used as part of security awareness campaigns [15]. For example, Denning [16], [17] provides with Control-Alt-Hack a game to raise security awareness by letting players become white hat hackers. Control-Alt-Hack does not focus on threat analysis or security requirements elicitation, but rather places emphasis on awareness. Therefore, it is set in a fictional scenario. In addition, the players use attacks that are predefined on the cards and do not need to elicit attacks on their own. The reason is the aim of awareness, which limits the game to increasing its players' knowledge about the existence and potential harm of hacking attacks.

We believe that there is a major benefit from eliciting security requirements using employees of a company in such a game for social engineering. In contrast to security engineers, common employers have the benefit of knowing their daily routine well. Namely, they are aware of business processes and their contexts, and especially deviations from provisions. Additionally, they know about their (and their co-workers') security knowledge, attitudes towards security rules and policies, and past behavior. In short, the employees are unconsciously aware of the human vulnerabilities in a company.

We propose to use a game (see Figs. 1 and 2) to make these threats explicit, which lets them play the role of a social engineering attacker. The game provides the required information about human behavior patterns such as the herd principle (if everyone is doing it, I do it as well) and attack scenarios that social engineers use such as phishing.

In order to provide the validity of these principles and attack scenarios, we took all of them from scientific publications. The game enables employees to learn about social engineering, while practicing immediately. This immediate application of learned knowledge has proven to have lasting effects [18].

The game works as follows. Employees propose social engineering threats and the other players rate their validity based on their knowledge of the context, e.g. employee Anton would fall for a phishing mail only if he is under time pressure for a deadline. This leads to a ranking of the proposed threats. Afterwards the threats are the basis for security requirements that shall prevent them.

Currently companies focus on two options for addressing the social engineering problem.

Figure 1: The Cards of our Game

Firstly, companies can conduct *security awareness trainings* in which employees are told about the threat of social engineering. These trainings are often mandatory for employees and don't have a lasting effect[4]. As a cheaper variant, security awareness campaigns try to achieve the same goal, but face the same problems than trainings. In general, they are not well adapted to the employees' weaknesses.

Secondly, companies hire penetration testing companies that *attack* their clients and show weaknesses.These kind of penetrations tests are rarely done, because they come with a number of problems, e.g. a lot of effort needs to be invested beforehand to address legal issues [19]. At best, when those penetration tests are conducted, the tester finds flaws and companies can educate the affected employees. However, experiments have shown that these approaches are difficult, because humans are easily demotivated when confronted with the results [20].

We propose to solve this issue by playing our serious game for social engineering threat analysis. Our target audience consists of all employees of a company, security aware IT administrators and security engineers, as well as secretaries or sales persons. The reason why we even want security aware employees to play is, that these usually focus on technical threats and have currently little to no support for eliciting social engineering support.

The remainder of our paper is organized as follows. Section II reports on the goals of our project. Section III provides an overview of serious games in particular with regards to security and security requirements engineering approaches. Section IV describes the game and its design process. Sect. V reports on our evaluation of the game and shows resulting threats and security requirements. Section VI concludes and provides directions for future work.

II. Project Goals

A. Goals

As motivated by Sect. I our main goal is to provide structured means to elicit and prioritize social engineering security requirements. This includes:

[4]https://citadel-information.com/wp-content/uploads/2010/12/Beyond-Awareness-Training-Its-Time-to-Change-the-Culture-Stahl-0504.pdf

Figure 2: Our Serious Game for Social Engineering

- Considering a context specific to a company that shall be protected, which means considering personal traits of its employees, weaknesses in its processes, and lack of awareness or even misguided security attitudes and policies. If we do not provide essential support for **context-specific** threats the players run the risk to come up with generic and meaningless threats. This would be fine for raising awareness, but not for threat elicitation.
- Basing our game on **existing research**, which has been thoroughly evaluated by international researchers in the field of social engineering. We wanted to avoid bias by making up social engineering elements (behaviors and attack scenarios) by ourselves or external consultants and missing relevant fundamental elements.
- Keeping our game **simple** allows the players to focus on the threat analysis and spend as little effort as possible on learning and following the game's rules. This allows them to focus most of their cognitive powers on eliciting the threats.
- Making the game **entertaining** is of utmost importance. According to Klimmt [21, p. 256f] enjoyment during the game generates attention and interest. An external security engineer would need to understand the company (processes, policies, employees' capabilities and attitudes) and get domain knowledge in order to elicit threats. We believe it is easier and more cost-effective to train the people that know the context of their work really well in threat analysis. The highest danger of the participation of non security experts is the *looking out of the window*[5] effect, which describes the participants' boredom leads them to stop participating and spend their time looking out the window and thinking of other topics. Our aim is to avoid this effect by engaging the players in an enjoyable experience.

[5]This effect was introduced to the authors by Ketil Stølen.

B. Why a Game?

This section is mainly based on the argumentation of Denning [16] for their security awareness game. We extended Denning's argumentation with arguments from research on serious games. As a result, we believe that a serious game can fulfill our project goals. If designed properly, a serious game can be an appropriate tool for supporting context-specific threat analysis to different kinds of employees. In short:

- Games can be fun, which gets employees involved.
- Games provide a realm that encourages employees to be creative and try new ways of thinking
- Games are intended to be engaging and entertaining, which gets employees to play again and again.
- A game provides a realistic scenario, but the players do not need to fear consequences, because "playful action [...] is intentionally limited to a situational frame that blocks out further consequences of action results."(cf. Klimmt [21, p. 253]) Klimmt points out, that direct consequences are a reduction of complexity, because players do not even need to think about consequences. Another consequence is the accessibility of imagined contexts and activities; fantasy allows role-play in contexts that would not be feasible, appropriate or desirable otherwise. This mind-set exactly matches our aim to make players think like an attacker.

We could have designed this game as a computer game. Both formats have their benefits and limitations. We decided to design a physical tabletop game mainly, because the social setting of the game involves the physical presence of potential victims and the players are reminded of their vulnerabilities while playing. These victims or people that know them well can participate in the discussions about threats and may be reminded of their actual behavior by their presence. Furthermore, Denning's reasons apply in our case, as well.

- Physical games may be attractive to people who dislike computer games.
- Physical games require no hardware or digital resources, except for a table.
- Physical games allow to browse its components such as principles without playing.

C. Target Audience

When designing the game, we had to consider the trade-off between designing a very general and generic game and one specific for a certain target group. While a game appealing to as many people as possible may be broadly applicable, a more targeted version may benefit from domain knowledge and may be more helpful for the players. We decided to target the middle and design a game for employees without consideration of properties specific for certain industrial sectors.

a) Primary Audience: Our game addresses *company employees* that work with computers and information assets. In particular, we want to engage *security engineers* and *IT administrators* in social engineering threat analysis. We claim that these have initial security knowledge which makes it

easier for them to get introduced to the topic. On the other hand the human engagement necessary when dealing with social engineering is fairly new to many of these population and our game shall help with this task.

b) Secondary Audience: Persons in a company that work with information assets are the entire *Administration* staff. We welcome their engagement in the game in order to be motivated and encouraged to tackle social engineering. Ideally we mix this second audience with the first, so that knowledgeable security people can explain security concepts and procedures during the discussions of the game.

In the future, we plan to provide introductionary material and further examples to make the game appealing to a broader population.

III. Background and Related Work

We are not aware of a serious game for social engineering to elicit security requirements. We report on the following works relating to serious games in software planning and security engineering.

Serious games have demonstrated a significant potential in industrial education and training disciplines [22], [23], [24], given that organizations care for players' privacy and working atmosphere and especially do not use gaming data for appraisal or selection purposes, and clearly communicate this to employees [25].

In particular, games for IT security preparedness in the electricity industry in Norway [26] have helped to determine the right composition of response teams in terms of competencies. These exercises have the potential to optimize current emergency practices and they offer the possibility to evaluate new practices in a realistic setting.

The *planning poker* game [27], [28] provides a collaborative method for estimating efforts for software engineering. The players take turns to estimate the efforts of a task in the first round, discuss the reasoning for their estimations and estimate again in a second round. The results are well agreed upon resource estimates. The variant of planning poker for software security called *protection poker* [11], [12] provides a way for understanding and prioritizing risks. The game lets software engineers estimate the value of assets and the potential damage of threats towards these assets. The players suggest and discuss estimates for these values similar to planning poker. Finally the players quantify the risk for each asset and threat pair by multiplying their values. These pairs are placed in descending order by their risk values, which results in a prioritized list of risks. The game has also the benefit that software engineers have a simple way to discuss and learn about security concerns and measures. The authors found reasonable indication for this statement based on their empirical evaluations with academics [11] and practitioners [12]. In contrast to our work, this game does not use cards, but estimates on paper or boards and does not focus on social engineering. In the future, we can combine our games as follows. Our threats can be input for protection poker, which adds risk assessment to our threats.

Shostack [13], [14] argues as well that teaching software engineers about security is more favorable than using security engineers to conduct the threat analysis, because security engineers have to invest a lot of time to understand the work of the software engineers. This understanding is essential to discover vulnerabilities. In contrast, software engineers are more familiar with possible vulnerabilities of their systems, if they are taught about threat analysis. Thus, the author developed a card game called *Elevation of Privileges*. In contrast to the games described before it is a physical card game[6]. Each player draws several threats. In turns, the players then explain how these threats could manifest with regard to the software they are currently engineering. If a player can convince the other players that her threat is worth a bug investigation, a request for an additional feature or even a design change, she gets a point. The player with the most points by the end of the game wins. In contrast to our work, Shostack focuses on software security and software engineers as a target audience, while our game is for any kind of employees that work with information assets.

Games are also effective in security awareness campaigns [15], which aim to make people aware of IT security threats. The serious game Control-Alt-Hack from Denning [16], [17] is a tabletop game that lets players take the role of managers of a security penetration testing company. The company attacks its customers with their consent and the player that achieves the most successful attacks and earns subsequently the most money wins. The success of the attacks is decided by a roll of the dice. The players learn about existing attacks and the damage they can cause within the fictional setting. In contrast to our work, the game has a focus on awareness, and therefore no context-specific threats are elicited or security requirements documented.

The security cards[7] is a deck of cards that contains cards of the types impact on humans, adversary motives, adversary resources, and adversary methods. The aim of this game is to brainstorm about threats. In contrast to our work these cards do not come with a clear set of rules and are not based on literature, but are more vague. For example, an adversary's method is processes and asks the players to come up with a bureaucratic process for an attack. This level of abstraction provides less guidance than our card games.

Further available games are [d0x3d!][8] a tabletop game designed to raise awareness to network security terminology and attacker models. The card game *Exploit!*[9] is an entertainment game for security engineers. OWASP Cornucopia[10] trains threat modeling and risk assessment for software applications. However, none of these games addresses social engineering threat elicitation with employees.

[6]The Elevation of Privileges (EoP) Card Game: https://www.microsoft.com/en-us/SDL/adopt/eop.aspx

[7]The security cards: http://securitycards.cs.washington.edu

[8]The [d0x3d!] game: http://www.d0x3d.com

[9]Core Impact: Exploit! http://www.coresecurity.com

[10]Cornucopia https://www.owasp.org/index.php/OWASP_Cornucopia

Capture-the-Flag[11] games make the players compete in simulated security attacks. These have been extended to the realm of social engineering[12]. These competitions select social engineers that attack existing companies, but these are not employees of these companies and limit themselves to telephone based attacks. The companies are informed of the results, but these often do not volunteer to be attacked and as unwilling participants the positive impact these studies can have seem rather limited. In any case, Social Engineering Capture the Flag are more a kind of social penetration testing than threat elicitation. From the companies' perspective, they therefore come with the problems already mentioned in Sect. I.

IV. Designing the Game

We could not identify a game that provides structured context-specific threat analysis, is based on existing research, is simple and engaging (see Sect. III). Thus, we decided to create our own game mechanics and improved them over a number of feedback rounds. Our game on social engineering consists of three sections: *Preparation* the game considering the players' context, *Playing* the game and eliciting threats and *Debriefing* the players including prioritizing threats. Each of the sections may consist of several phases. In this Section, we present the game rules along with our design rationales.

Section 1: Preparation

1. Create an Overview Plan Provide an *overview plan* of the department by using the fire escape plan. This plan has to be augmented with the assets of the company, the people working in that department, and their locations, as well as communication channels e.g. VoIP, Email, etc (cf. Fig. 3). All players should be involved in the creation or have to check the plan for completeness.

Reasoning: We base this step on the fire escape plan of the department, because it is easily available since it often is publicly hung out to show escape routes. Additionally, the plan shows fire-extinguishers, fire alarm buttons, and escape ways, which may be used by the players in an attack. Lastly, the natural consequence of the players checking it for flaws is that they are familiar with it at the beginning of the game and further discussion in the game is focused on the attacks and not the setup.

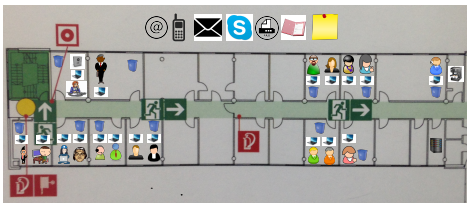


Figure 3: Overview Plan

[11] Capture the Flag: https://www.defcon.org/html/links/dc-ctf.html
[12] Social Engineering CTF: http://www.social-engineer.org/category/ctf/

Section 2: Playing

In the Playing section, the players take the role of the attacker. It consists of the following phases:

1. Draw Human Behavioral Pattern Card Each player draws a card from the deck of *human behavioral patterns* (principles). Users behaving according to one of the principles can be exploited by social engineers. One example for the patterns is the so-called *Need and Greed principle* that states "Your needs and desires make you vulnerable. Once hustlers know what you really want, they can easily manipulate you. "A sample card is shown in Fig. 4.

Reasoning: The human behavioral patterns are based on the work of Stajano and Wilson [29], who describe why attacks on scam victims may succeed. We extended the set of behavioral patterns[13] by patterns found in work on social engineering from Gulati [30] and Peltier [31].

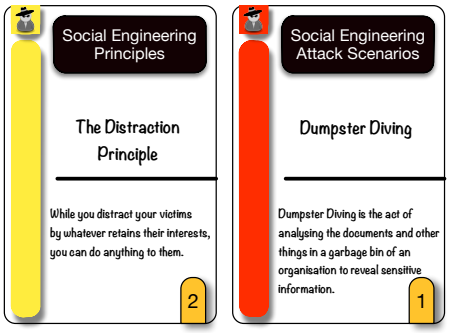


Figure 4: Principle (Left), Attack Technique (Right)

2. Draw Attack Scenario Cards The next step is that each player draws three cards from the deck of the *social engineering attack techniques* (scenarios). For example, reverse social engineering comprises creating a problem for the selected person and solving it for him. The gained trust is used to ask the victim for a favor. A sample card is shown in Fig. 4.

Reasoning: The used attack techniques are mostly based on the work of Krombholz et al. [32]. We also extended the set of attack techniques[14] adapted from the work of Gulati [30], Peltier [31], and Chitrey et al. [33]. Since most attacks are only related to a subset of the behavioral patterns in an appropriate manner, we allow the players to take three cards.

3. Choose Attacker Type Each player gets one *attacker type* card. The card has two sides shown in Fig. 5. One for an inside attacker, who is a well known member of the organization. And one for an outside attacker, who is unknown to the members of the organization.

[13] A full list of all human behavioral patterns along with the corresponding reference may be retrieved on http://pape.science/social-engineering/.
[14] A full list of all attack techniques along with the corresponding reference may be retrieved on http://pape.science/social-engineering/.

Reasoning: Insiders have already established trust in the organizations, which leads to an easier starting point for an attack. Outsiders have to establish trust in the organization first before conducting the attack. The players should think about what kind of attacker they are and plan their attack accordingly. For example, an insider might need to cover his tracks more carefully or pass the buck to co-workers while an outsider has to provide a reason for being in the building.

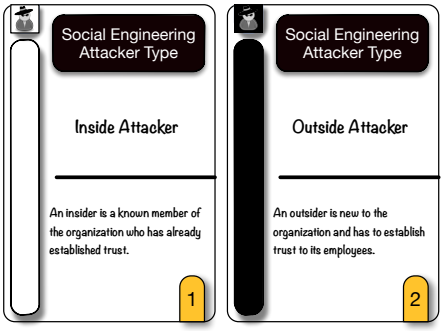


Figure 5: Attacker Card (Front- and Backside)

4. Brainstorming In the brainstorming phase the players take the role of the attacker. Each player thinks of how to conduct one of the three attacks to exploit the behavioral pattern of an employee. The exploit targets one person in the overview diagram and an asset. Moreover, the player has to choose if she is an insider or outsider of the organization. The players get five minutes to think about and elaborate their attacks.

Reasoning: We experimented with different time frames for the brainstorming in feedback sessions and discovered in 4 sessions (n=3,n=2,n=4,n=4) that players need on average between 4 to 5 minutes to elicit a threat. A too short time frame showed to annoy the players while too much time got them easily distracted.

Rounds of the game: Each player proposes an attack in the fashion explained below. This iterates until all persons iterated at least twice. We denote a *turn* as one player presenting an attack along with the discussion and getting points. A *round* consists of turns of all players. After each round, the players restock their cards. The brainstorming phase in the iterations may be shortened as needed by the players.

5. Attack The active player presents his attack to the group. Each attack consists of a principle, an attack scenario, an attacker, a victim, a communication channel and a targeted asset (c.f. Tab. I). Note that once a player has proposed an attack it is finalized and cannot be changed anymore by the player.

Reasoning: The players should finalize their attack, because otherwise the players could always adapt their attack to address any concerns that may arise and gain full points. While this has still lead to lively discussions, it showed that players were dissatisfied because the awarding of points did not reflect the players effort. As discussed in Sect. II-B it is important to retain the playful character of the players' actions.

6. Discussion The discussion starts with a feasibility reasoning of the proposed threat. The players discuss first, if the attack is feasible, in which case the player gets 2 points. If the player received help when describing the attack or the attack is plausible, but infeasible (e.g. because the attacked person has a special training to resists the described attack), the player gets 1 point. If the proposed attack is not plausible the turn ends immediately and the player gets no points.

In case the player received more than one point, a compliance discussion follows. *Principle:* If the attack described by the player is a perfect match, the player gets 2 points, if it matches only somehow, he gets 1 point. *Scenario:* If the attack described matches the presented attack technique card, the player scores 1 point. *Attacker:* Finally, the players discuss if the inside attacker (1 point) and outside attackers (2 points) card matches the attacker type in the proposed threat.

Reasoning: We first want to establish if the attack is intuitively working in the minds of the players or if reasonable doubts exist. If the doubts are so strong that no players believes this attack can work we have a punishment installed in the game (0 points and end of turn). Afterwards, we would like to reward the players to think about the behaviors and attack scenarios on their cards, as well as the different approaches of inside and outside attackers.

7. Improvements In addition, the other players can also propose improved versions of an attack and gain 2 points for a major improvement or 1 point for a minor improvement. The points are granted by the other players.

Reasoning: We want to get the other players variations of the threat in order to explore their variations. Any missing threat during a security analysis presents a risk that is not considered and subsequently not protected against.

Section 3: Debriefing

In the debriefing phase, the players reflect their attacks. They may be supported by the company's security personnel.

Prioritize Threats We propose the following activities: (1) identify the most relevant threats of social engineers in your organization (e.g. based on likelihood to succeed and damage they potentially cause), (2) try to figure out why some people were attacked more often and (maybe) others not at all, (3) analyze why some communication channels were used more often than others, and (4) determine which assets were attacked more often than others.

Reasoning We aim to foster discussions about how severe social engineering attacks can be for an organization and find out which are the main security concerns for social engineering respectively.

Document Security Requirements We use a similar approach than Misuse Cases [34] to map threats to security requirements that specify the underlying security problem.

Reasoning We want this step to be simple and based on some well established work. The misuse case fulfills that criteria.

V. Evaluation

A. Sampling

We evaluated the game in practical experiments at the University Frankfurt and the Technical University Munich. We played the game with 27 players that are full time employed and 3 senior researchers have participated in the game in the role of a game master. The distribution of the players is the following (see Fig. 6): 5 players are employed at the University Frankfurt, 19 at the Technical University Munich and 1 is employed at a telecommunications company. Among the players were 2 senior researchers, 19 researchers, 4 members of the IT administration staff, 1 secretary and 1 professor. In particular, the players held masters' degrees in computer science (18), business information systems (4), economics (1), and IT security (1). In addition, 3 players have a PhD in computer science, while 4 players do not have academic degrees (see Fig. 7). We did not use students in our evaluations, but scientific employees and members of the administrative staff. The reason for this is that the target audience of the game consists of company employees and we identified a sample set that reflects our target audience.

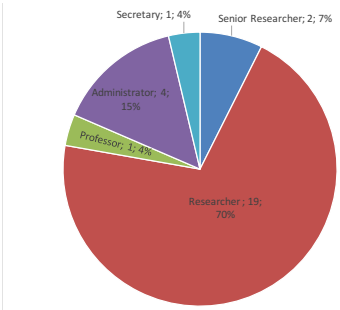


Figure 6: Player's Professions

B. Operation

We played the game in 7 individual sessions with 3 to 4 players and 1 game master in each round. In total, 49 turns of the game were played and 17 hours of playing time. Note that the time of playing the game varied depending on the length of the discussions of the feasibility of a proposed attack.

For the first two sessions we introduced social engineering and the rules of the game in a 15 minute presentation. Afterwards, we decided to shorten the introduction in order to get the players involved with the game sooner. Hence, sessions 3 to 7 are just introduced with a 5 minute introduction. We

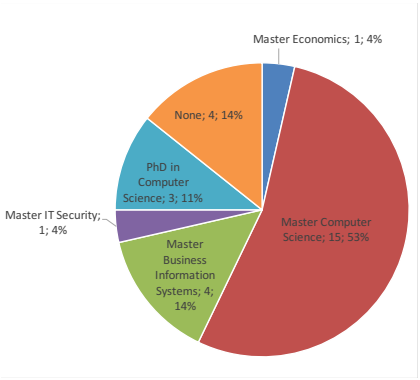


Figure 7: Player's Academic Degrees

devised a handout[15]. for the players in order to gain easy access to the rules at any time and handed it to them before starting a round. We played the game in a closed room so the players would not be distracted by any outside influence. Some of the players mentioned that our instructions on an A4 paper are too long for reading while playing. Therefore, we also provided a short version of the rules.

The game masters initiated the game with issuing the cards and just motivated the participants to elicit and discuss social engineering threats. They ensured that everyone's opinion is heard. The game masters did not voice any opinions during the discussions, they just documented the choices regarding points of the participants. Afterwards the game masters conducted the debriefing of the players of the game. The results of the debriefing is reported as part of the data analysis (c.f. Tab. I).

C. Data Analysis

We present the resulting statistics of the game in the following. We played 49 turns of the game resulting in 33 plausible attacks, 9 feasible attacks, and 7 non feasible attacks (see Fig. 8). Hence, the majority of the elicited attacks were plausible. Note that the following statistics focus only on the 42 plausible and feasible attacks. We exclude the non feasible attacks for the following analysis. We categorized the victims attacked in our threats to the following types: secretary, employee, and IT administrator (see Fig. 9). Our analysis revealed that employees are the most often attacked victims. We assumed before playing the game that this would be the secretaries, because they are assumed to be the weakest spot. The reason for this is that they have the least amount of university education and the most contact with people. Moreover, they are doorkeepers of the department heads and

[15]The handout is available at: http://pape.science/social-engineering/

Table I: Social Engineering Threats Elicited during our Evaluation

Nr.	Context Knowledge	Attack	Asset	Principle	Attack Scenario
1	Tim is seeking for attention and likes to be admired for his achievements.	A member of an intranet security discussion board invites Tim to participate in an honorary event and asks the Tim to log in with his credentials to the intranet side using a specific link.	Credentials	Distraction	Waterholing
2	Jim flies to the United States from Germany with Lufthansa and they just announced a strike. Jim is watching his email closely to get any information about delays quick and deal with them.	The attacker fakes a Lufthansa email with an updated travel itinerary and attaches some malware to this email. The malware would gain him access to the Jim's PC and all digital assets on it.	Notebook Data	Time Pressure	Mail Attachment
3	Bob is using Yahoo Mail, which forces him to re-enter his credentials after 2 weeks continuously being logged in.	Bob proposes to attack himself using the outlined weakness in Yahoo Mail. If an attacker would fake the popup, he would probably (re-)enter his credentials	Email Data	Ignorance and Carelessness	Popup Window
4	Steve always leaves his office door and computer unlocked. The cleaning guy is quite dominant when cleaning the rooms.	An attacker can just enter his office pretending to be a (new) cleaning guy, so he can just enter and send an email using his computer and open an attachment with a Trojan.	Notebook Data	Laziness	Support Staff
5	Robert's family is about to arrive in the city to celebrate his PhD submission. He also is printing his Phd-thesis at the moment. Robert gets a call from his family who arrives by train.	The attacker would be around and offer him to finish copying his dissertation. Due to Robert's stress with his dissertation and family arriving he would welcome help. The attacker would then steal data from his dissertation.	Dissertation	Trust Principle	Direct Approach
6	Claudia is a new employee and worried about her reputation. She is using the local WiFi access and the company is communicating with a chat tool.	The attacker would send her some links that turn out to be pornography in the chat tool, after that the attacker will call her and pretend to be a system administrator and pressure her to reveal confidential information for not letting anybody know about the pornography.	Confidential information	Trust Principle	Direct Approach
7	Bernhard needs a lot of computational power to run experiments. He does not have sufficient resources and a tight deadline to deliver results. He just ordered more IT resources.	The attacker spoofs the email of the IT administration and sends him an email pretending to be the administration. The email asks to open an email attachment that contains a new form he has to fill in if he wants to get the resources he previously ordered. The attachment contains a malware.	PC data	Need and Greet	Email Attachment
8	Jean has to work a lot with the financial administration due to project billing issues for a European research project she is working on.	The attacker pretends to be from the finance administration and gain her credentials for the website the European Project is used for billing. By telling that some issues need to be resolved and proposing to take care of them for her she would gladly give the attacker her credentials.	Credentials	Guilt (No points)	Direct Approach
9	Torben googles himself regularly to check his reputation in the web.	The attacker prepares a site with information about him and with exploits. The attackers would try to get it in the google ranking and wait for him to google himself. If he checks the results and notices the new page, he'll browse it.	PC data	Guilt (No points)	Direct Approach (No points)
10	Recently, there has been a bomb threat and the administration asked everyone to leave the building for "technical reasons". Further information was promised the next days.	Impersonate someone from the health department and claim that all people have to leave the building due to recently discovered asbestos or start a fake fire alarm to access the boss's office for a couple of minutes.	Data in Office	Fear of the Unknown	Third Party Authorization

often hold a lot of access privileges. However, this assumption turned out to be wrong. Moreover, we did not expect more than 10% of the attacks directed towards the IT administration, because these are supposedly the most well trained employees with regard to social engineering. Furthermore, we present the distribution of attacks towards employees in detail in Fig. 10 right. The blue employees are secretaries, the green ones are administrators and the red ones are scientific employees. The number following the name is the number of times that person was attacked. All of the names are pseudonyms for real people. The person that suffered the most attacks is Monja a secretary with overall 8 attacks. In contrast, all other victims suffered between 1 and 3 attacks.

The ratio between insider and outsider attackers is 22 outsider attacks to 20 insider attacks (see Fig. 8). We expected a large ratio of insider attacks, because these are easier to elicit, due to the fact that inside attackers have already established trust. In particular, the players can attack as themselves. However, the statistics show that these numbers are almost even and we could not reveal a significant preference for either attacker type.

Table II: From Threats to Requirements

Nr.	Threat	Security Requirement
1	A member of an intranet security discussion board invites Tim to participate in an honorary event and asks the boss to log in with his credentials to the intranet side using a specific link.	A security awareness training has to teach Tim and other employees to investigate links from unknown sources, even when under time pressure. These investigations can be delegated, e.g., to the IT security team.

We present an excerpt of the threats we elicited in our evaluation in Tab. I and show the domain knowledge these contain. The table outlines the drawn cards and targeted assets, as well. Note that even if these attacks are plausible, in some cases the players did not receive the points for principles or attacks, because her attack did not match the received cards.

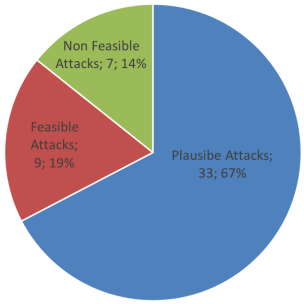


Figure 8: Attack Rating

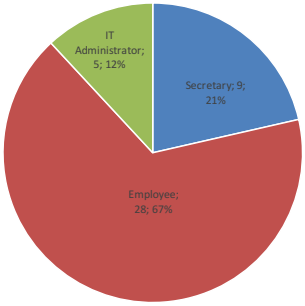


Figure 9: Victim Type

The final step of our approach is to formulate security requirements based on these threats. We provide an exemplary threat and requirement pair in Tab. II. The requirements shall contain a constructive procedure to support the possible victims in evading the elicited threats. In the future, we will look into how to do a reconciliation of multiple social engineering security requirements to derive an entire awareness training from it.

In addition, we deem it important that the employees understand that they will not be punished if they fell for a social engineering attack, but limit the damage in informing the security incident management team. The person that does this debriefing has to ensure that employees understand that they can resist this attack with proper training and motivate them to do so. The understanding of the social engineering attacker due to the precise attack presented should help employees to gain confidence that they can adopt a resistance strategy.

VI. Conclusion and Future Work

Social engineering attacks are a significant problem for IT security. Even for IT security professionals it is challenging to elicit security requirements for social engineering threats.

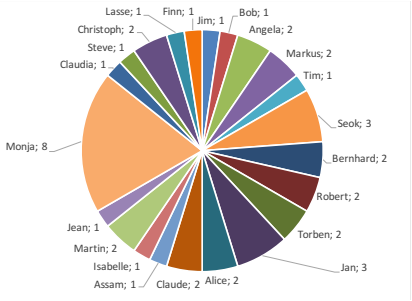


Figure 10: Victim Type Detail

Commonly, social engineering threat assessment involves penetration testers that execute attacks on their customers and report security requirements to them afterwards. This involves the deception of people and a possible violation of their privacy rights and provides only a small fraction of all attack vectors. We propose an alternative to these techniques that does not involve the lying to people, does not require external security consultants, carries less risk of privacy violations, and utilizes domain knowledge of the employees of these companies. These employees have due to their work experience in the company the most relevant information to assess social engineering vulnerabilities in themselves and their colleagues.

Our proposed solution utilizes a card game that employees of a company play to elicit social engineering threats and subsequent security requirements. These know the domain well and learn about social engineering in a structured way while playing the game. Security consultants are more familiar with social engineering, but they have to learn about the domain to elicit relevant context-specific threats. We argue that employees can be taught this knowledge with our game and at least contribute to the threat analysis and security requirements elicitation effort.

Our main contributions are listed in the following:

- Employees learn about different facets of social engineering acts e.g. how social engineering attacks are composed. They learn by applying the knowledge when becoming an attacker in the game. The learning and application of social engineering while having fun playing creates lasting knowledge on the subject.
- The domain knowledge of the players and in particular their observations during their daily work allows them to elicit context-specific attacks.
- The plausibility of the proposed attacks are rated by the employees, again by applying valuable insights of the domain in their argumentation. Hence, not plausible attacks in this specific context are eliminated quickly. It also leads to a prioritization of threats and their respective security requirements into plausible attacks and only feasible ones.

- The employees are warned about threats they may face in their daily lives and develop a sense of suspicion when being attacked. Threats being elicited with our game have domain specific information, which makes them realistic.

In the future, we are planning to create a context-independent version of the game that can be used without preparation in security awareness campaigns.

VII. Acknowledgements

We thank all the players of our game that provided us with invaluable feedback and spend their precious time with us improving the game.

This research has been partially supported by the Federal Ministry of Education and Research Germany (BMBF) via the project SIDATE (grant number 16KIS0240) and the TUM Living Lab Connected Mobility (TUM LLCM) project funded by the Bayerisches Staatsministerium für Wirtschaft und Medien, Energie und Technologie (StMWi).

References

[1] A. van Lamsweerde and E. Letier, "Handling obstacles in goal-oriented requirements engineering," *IEEE Trans. Softw. Eng.*, vol. 26, no. 10, pp. 978–1005, Oct. 2000. [Online]. Available: http://dx.doi.org/10.1109/32.879820

[2] H. Mouratidis and P. Giorgini, "Secure tropos: A security-oriented extension of the tropos methodology." *Journal of Autonomous Agents and Mult-Agent Systems*, 2005.

[3] L. Liu, E. Yu, and J. Mylopoulos, "Security and privacy requirements analysis within a social setting," in *Proceedings of the 11th IEEE International Conference on Requirements Engineering*, ser. RE '03. Washington, DC, USA: IEEE Computer Society, 2003, pp. 151–. [Online]. Available: http://dl.acm.org/citation.cfm?id=942807.943910

[4] P. Herrmann and G. Herrmann, "Security requirement analysis of business processes," *Electronic Commerce Research*, vol. 6, no. 3, pp. 305–335, 2006.

[5] T. Li and J. Horkoff, *Advanced Information Systems Engineering: 26th International Conference, CAiSE 2014, Thessaloniki, Greece, June 16-20, 2014. Proceedings*. Cham: Springer International Publishing, 2014, ch. Dealing with Security Requirements for Socio-Technical Systems: A Holistic Approach, pp. 285–300.

[6] T. Li, E. Paja, J. Mylopoulos, J. Horkoff, and K. Beckers, "Holistic security requirements analysis: An attacker's perspective," in *Requirements Engineering Conference (RE), 2015 IEEE 23rd International*, 2015, pp. 282–283.

[7] T. Li, J. Horkoff, E. Paja, K. Beckers, and J. Mylopoulos, *The Practice of Enterprise Modeling: 8th IFIP WG 8.1. Working Conference, PoEM 2015, Valencia, Spain, November 10-12, 2015, Proceedings*. Springer International Publishing, 2015, ch. Analyzing Attack Strategies Through Anti-goal Refinement, pp. 75–90.

[8] S. H. Houmb, S. Islam, E. Knauss, J. Jürjens, and K. Schneider, "Eliciting security requirements and tracing them to design: An integration of common criteria, heuristics, and umlsec," *Requir. Eng.*, vol. 15, no. 1, pp. 63–93, 2010.

[9] A. Herrmann, A. Morali, S. Etalle, and R. Wieringa, "Riskrep: Risk-based security requirements elicitation and prioritization," in *1st International Workshop on Alignment of Business Process and Security Modelling, ABPSM 2011*, ser. Lecture Notes in Business Information Processing. Berlin, Germany: Springer Verlag, October 2011, pp. 1–8. [Online]. Available: http://doc.utwente.nl/78045/

[10] D. Ionita, J. W. Bullee, and R. J. Wieringa, "Argumentation-based security requirements elicitation: The next round," in *Evolving Security and Privacy Requirements Engineering (ESPRE), 2014 IEEE 1st Workshop on*, Aug 2014, pp. 7–12.

[11] L. Williams, M. Gegick, and A. Meneely, "Protection poker: Structuring software security risk assessment and knowledge transfer," in *Proceedings of International Symposium on Engineering Secure Software and Systems*. Springer, 2009, pp. 122–134.

[12] L. Williams, A. Meneely, and G. Shipley, "Protection poker: The new software security "game"," *Security Privacy, IEEE*, vol. 8, no. 3, pp. 14–20, May 2010.

[13] A. Shostack, *Threat Modeling: Designing for Security*, 1st ed. John Wiley & Sons Inc., 2014.

[14] ——, "Elevation of privilege: Drawing developers into threat modeling," Microsoft, Redmond, U.S., Tech. Rep., 2012, http://download.microsoft.com/download/F/A/E/FAE1434F-6D22-4581-9804-8B60C04354E4/EoP_Whitepaper.pdf.

[15] M. Gondree, Z. N. J. Peterson, and T. Denning, "Security through play," *IEEE Security and Privacy*, vol. 11, no. 3, pp. 64–67, 2013.

[16] T. Denning, A. Lerner, A. Shostack, and T. Kohno, "Control-alt-hack: The design and evaluation of a card game for computer security awareness and education," in *Proceedings of the 2013 ACM SIGSAC Conference on Computer & Communications Security*, ser. CCS '13. ACM, 2013, pp. 915–928.

[17] T. Denning, T. Kohno, and A. Shostack, "Control-alt-hack™: A card game for computer security outreach and education (abstract only)," in *Proceeding of the 44th ACM Technical Symposium on Computer Science Education*, ser. SIGCSE '13. ACM, 2013, pp. 729–729.

[18] F. L. Greitzer, O. A. Kuchar, and K. Huston, "Cognitive science implications for enhancing training effectiveness in a serious gaming context," *J. Educ. Resour. Comput.*, vol. 7, no. 3, 2007.

[19] G. Watson, A. Mason, and R. Ackroyd, *Social Engineering Penetration Testing: Executing Social Engineering Pen Tests, Assessments and Defense*. Syngress, 2011.

[20] T. Dimkov, A. van Cleeff, W. Pieters, and P. Hartel, "Two methodologies for physical penetration testing using social engineering," in *Proceedings of the 26th Annual Computer Security Applications Conference*, ser. ACSAC '10. ACM, 2010, pp. 399–408.

[21] C. Klimmt, "Serious games and social change: Why they (should) work," in *Serious games: Mechanisms and effects*, U. Ritterfeld, M. Cody, and P. Vorderer, Eds. Routledge, 2009.

[22] P. Petridis, K. Hadjicosta, V. S. Guang, I. Dunwell, T. Baines, A. Bigdeli, O. F. Bustinza, and V. Uren, "State-of-the-art in business games," *International Journal of Serious Games*, vol. 2, no. 1, pp. 55–69, 2015.

[23] J. Riedel and J. Hauge, "State of the art of serious games for business and industry," in *Proceedings of Concurrent Enterprising (ICE)*, 2011, pp. 1–8.

[24] C. E. Catalano, A. M. Luccini, and M. Mortara, "Best practices for an effective design and evaluation of serious games," *International Journal of Serious Games*, vol. 1, no. 1, pp. 12–25, 2014.

[25] M. Malheiros, C. Jennett, W. Seager, and M. A. Sasse, "Trusting to learn: Trust and privacy issues in serious games," in *Trust and Trustworthy Computing - 4th International Conference, TRUST 2011. Proceedings*, ser. Lecture Notes in Computer Science, vol. 6740, 2011, pp. 116–130.

[26] M. Line and N. Moe, "Understanding collaborative challenges in it security preparedness exercises," in *ICT Systems Security and Privacy Protection*, ser. IFIP Advances in Information and Communication Technology, H. Federrath and D. Gollmann, Eds. Springer International Publishing, 2015, vol. 455, pp. 311–324. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-18467-8_21

[27] J. Grenning, "Planning poker or how to avoid analysis paralysis while release planning," Object Mentor, Tech. Rep., 2002, https://renaissancesoftware.net/files/articles/PlanningPoker-v1.1.pdf.

[28] K. Molokken-Ostvold and N. Haugen, "Combining estimates with planning poker–an empirical study," in *Software Engineering Conference, 2007. ASWEC 2007. 18th Australian*, April 2007, pp. 349–358.

[29] F. Stajano and P. Wilson, "Understanding scam victims: Seven principles for systems security," *Commun. ACM*, vol. 54, no. 3, pp. 70–75, Mar. 2011. [Online]. Available: http://doi.acm.org/10.1145/1897852.1897872

[30] R. Gulati, "The threat of social engineering and your defense against it," *SANS Reading Room*, 2003.

[31] T. R. Peltier, "Social engineering: Concepts and solutions," *Information Systems Security*, vol. 15, no. 5, pp. 13–21, 2006.

[32] K. Krombholz, H. Hobel, M. Huber, and E. Weippl, "Social engineering attacks on the knowledge worker," in *Proceedings of Security of Information and Networks*, ser. SIN '13. ACM, 2013, pp. 28–35.

[33] A. Chitrey, D. Singh, and V. Singh, "A comprehensive study of social engineering based attacks in india to develop a conceptual model," *International Journal of Information and Network Security (IJINS)*, vol. 1, no. 2, pp. 45–53, 2012.

[34] G. Sindre and A. L. Opdahl, "Eliciting security requirements with misuse cases," *Requir. Eng.*, vol. 10, no. 1, pp. 34–44, 2005.

A.2 HATCH: Hack And Trick Capricious Humans – A Serious Game on Social Engineering

Kristian Beckers, Sebastian Pape, and Veronika Fries. HATCH: Hack and trick capricious humans – a serious game on social engineering. In *Proceedings of the 2016 British HCI Conference, Bournemouth, United Kingdom, July 11-15, 2016*, 2016. URL https://ewic.bcs.org/content/ConWebDoc/56973

HATCH: Hack And Trick Capricious Humans – A Serious Game on Social Engineering

Kristian Beckers
Technische Universität München (TUM)
Institute of Informatics
Boltzmannstr. 3
85748 Garching, Germany
kristian.beckers@tum.de

Sebastian Pape
Goethe-University Frankfurt
Faculty of Economics
Theodor-W.-Adorno-Platz 4
60323 Frankfurt, Germany
sebastian.pape@m-chair.de

Veronika Fries
Technische Universität München (TUM)
Institute of Informatics
Boltzmannstr. 3
85748 Garching, Germany
veronika.fries@tum.de

Social engineering is the illicit acquisition of information about computer systems by primarily non-technical means. Although the technical security of most critical systems is usually being regarded in penetration tests, such systems remain highly vulnerable to attacks from social engineers that exploit human behavioural patterns to obtain information (e.g., phishing). To achieve resilience against these attacks, we need to train people to teach them how these attacks work and how to detect them. We propose a serious game that helps players to understand how social engineering attackers work. The game can be played based on the real scenario in the company/department or based on a generic office scenario with personas that can be attacked. Our game trains people in realising social engineering attacks in an entertaining way, which shall cause a lasting learning effect.

Security, Methods, Education, Social Engineering, Serious Gaming

1. INTRODUCTION

Traditional penetration testing approaches often focus on vulnerabilities in network or software systems (Mitnick and Simon (2009)). Few approaches even consider the exploitation of humans via social engineering. While the amount of social engineering attacks and the damage they cause rises yearly the awareness of these attacks by employees remains low (Hadnagy (2010, 2016); Proofpoint (2016)).
Recently, serious games have built reputation for getting employees of companies involved in security activities in an enjoyable and sustainable way. While still preserving a playful character, serious games are used for e.g. security education and threat analysis (Williams et al. (2009, 2010), Shostack (2012, 2014), Denning et al. (2013)). We believe that there is a major benefit for adapting serious games specifically for social engineering (Beckers and Pape (2016a)). Our game aims at enabling common employees to elicit social engineering threats for their companies (real world scenario). Additionally, we have developed a generic scenario for training and awareness rising, which provides a description of a fictional office scenario with personas. In this paper we present our game, the generic scenario and our preliminary results of its application with students, academics, and industry.

Figure 1: *Picture of a Game Session*

2. DESIGN OF THE GAME

In short, the rules of the game are as follows:

1. Each player draws a card from the deck of *human behavioral patterns* (principles), e.g. the *Need and Greed principle*. The game is designed based on existing published work (e.g. Stajano and Wilson (2011), c.f. Beckers and Pape (2016b)).
2. Each player draws three cards from the deck of the *social engineering attack techniques* (scenarios), e.g. phishing. The game is

designed based on existing published work (e.g. Gulati (2003); Peltier (2006), c.f. Beckers and Pape (2016b)).

3. The players decide if they are insiders or outsiders to the organization.
4. Each player presents an attack to the group and the others discuss if the attack is feasible.
5. The players get points based on how viable their attack is and if the attack was compliant to the drawn cards. The player with the most points wins the game.
6. As debriefing, the perceived threats are discussed and the players reflect their attacks. They may be supported by the company's security personal.

3. INDEPENDENT SCENARIO

We created a generic scenario that people can relate to with little effort. We came up with the ACME office company, a medium sized producing company for paper. Therefore, we described 10 employees, their roles in the company, familiarisation with computers and attitudes towards security and privacy (see Fig. 2 as an example).

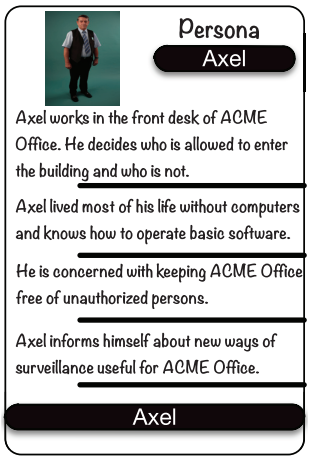


Figure 2: *A persona[1] within our ACME Office scenario*

4. PRELIMINARY RESULTS

To validate our research, we initially played the context-specific version with 25 full time employees of the Technical University Munich and Goethe-University Frankfurt with a university degree. We were initially interested if the players could elicit possible and context-specific threats for their respective environments. We played in total 49 turns of the game in which a player suggests a threat. The players deemed 42 of these threats possible and 7 were rated not possible by the players. The results suggest that the players were able to elicit threats with the game (c.f. Beckers and Pape (2016a)).

Afterwards, we were interested to measure if playing the game raises the security awareness of the players. Kruger and Kearny (Kruger and Kearney (2006)) measure security awareness in terms of knowledge (what an employee knows), attitude (what an employee thinks), and behaviour (what an employee does). We created a set of 14 questions that measured security awareness with relation to the attack scenarios in our game on a 5-point Likert scale. The answers range from *totally disagree* to *totally agree*. We assessed the questionnaires with games played with 10 full time employees from academia and 4 senior employees of an organisation A. The academics used our ACME office scenario and the senior employees the context-specific version of the game. We could measure on average between 0.5 and 1 point increase in security awareness with the players after they played HATCH. There was no statistical significant difference in persons who worked with ACME office scenario and the ones with the context-specific version of the game.

In future, we will try both versions of the game with a larger sample of participants and we are planning to measure the flow construct (Csikszentmihalyi (2000)) in relation to playing the game. In particular, we are planning to use the Flow Kurz Skala (Rheinberg et al. (2016)) to measure how intensive the player emerge in the game and correlate this to the difference in security awareness before and after the game. We assume that the flow experience is positively correlated to an increased security awareness. Additionally, we will create more generic scenarios to allow players with different background an easier access to the game.

5. ACKNOWLEDGEMENTS

We thank all the players of our game that provided us with invaluable feedback and spend their precious time with us improving the game. This research has been partially supported by Federal Ministry of Education and Research Germany (BMBF) within the focal point "IT-Security for Critical Infrastructures" (grant number 16KIS0240) and the TUM Living Lab Connected Mobility (TUM LLCM) project funded

[1] Picture is taken from Flickr https://flic.kr/p/Ch2gjk

by the Bayerisches Staatsministerium für Wirtschaft und Medien, Energie und Technologie (StMWi).

REFERENCES

Beckers, K. and S. Pape (2016a). A serious game for eliciting social engineering security requirements. In *Proceedings of the 24th IEEE International Conference on Requirements Engineering*, RE 16, pp. To Appear. IEEE Computer Society.

Beckers, K. and S. Pape (2016b). Theoretical foundation for: A serious game for social engineering. Technical report, Technical University Munich (TUM) and Goethe-University Frankfurt. http://pape.science/social-engineering/.

Csikszentmihalyi, M. (2000). *Beyond Boredom and Anxiety: Experiencing Flow in Work and Play* (25th Anniversary edition ed.). Jossey-Bass.

Denning, T., A. Lerner, A. Shostack, and T. Kohno (2013). Control-alt-hack: The design and evaluation of a card game for computer security awareness and education. In *Proceedings of the 2013 ACM SIGSAC Conference on Computer & Communications Security*, CCS '13, New York, NY, USA, pp. 915–928. ACM.

Gulati, R. (2003). The threat of social engineering and your defense against it. *SANS Reading Room*.

Hadnagy, C. (2010). *Social engineering: The art of human hacking*. Indianapolis: John Wiley & Sons.

Hadnagy, C. (2016). The social engineering infographic. Technical report, Social Engineer, Inc. http://www.social-engineer.org/social-engineering/social-engineering-infographic/.

Kruger, H. A. and W. D. Kearney (2006). A prototype for assessing information security awareness. *Comput. Secur. 25*(4), 289–296.

Mitnick, K. D. and W. L. Simon (2009). *The Art of Deception*. Wiley.

Peltier, T. R. (2006). Social engineering: Concepts and solutions. *Information Systems Security 15*(5), 13–21.

Proofpoint (2016). The human factor report 2016. https://www.proofpoint.com/us/human-factor-report-2016.

Rheinberg, F., R. Vollmeyer, and S. Engeser (2016). Flow kurz skala. Technical report. http://www.psych.uni-potsdam.de/people/rheinberg/messverfahren/FKS-englisch.pdf.

Shostack, A. (2012). Elevation of privilege: Drawing developers into threat modeling. Technical report, Microsoft, Redmond, U.S. http://download.microsoft.com/download/F/A/E/FAE1434F-6D22-4581-9804-8B60C04354E4/EoP_Whitepaper.pdf.

Shostack, A. (2014). *Threat Modeling: Designing for Security* (1st ed.). John Wiley & Sons Inc.

Stajano, F. and P. Wilson (2011, March). Understanding scam victims: Seven principles for systems security. *Commun. ACM 54*(3), 70–75.

Williams, L., M. Gegick, and A. Meneely (2009). Protection poker: Structuring software security risk assessment and knowledge transfer. In *Proceedings of International Symposium on Engineering Secure Software and Systems*, pp. 122–134. Springer.

Williams, L., A. Meneely, and G. Shipley (2010, May). Protection poker: The new software security "game". *Security Privacy, IEEE 8*(3), 14–20.

HATCH: Hack And Trick Capricious Humans
A Serious Game on Social Engineering

Kristian Beckers and Sebastian Pape and Veronika Fries
Technische Universität München
Goethe University Frankfurt

Technische Universität München

GOETHE UNIVERSITÄT FRANKFURT AM MAIN

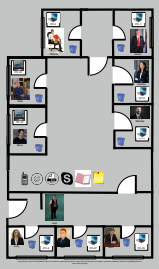
Objectives

A serious game on social engineering which aims to:

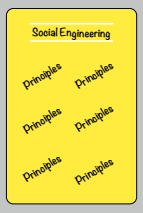
- Train the players on social engineering techniques
- Identify possible weaknesses to social engineering

Preparation

1. Present an *overview diagram* of the company that shall play the game. This diagram has to include the physical architecture of the company, the people working in that company and their locations, as well as communication channels e.g. VoIP, Email, etc. Finally the diagram has to show vital assets of the company, e.g., valuable information on a computer system. All players check the diagram for completeness and as a natural consequence should be familiar with it at the beginning of the game.

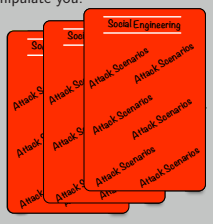


Draw Cards



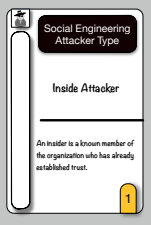
2. Each player draws 1 card from the set of *human behavioural patterns*.

   The card deck contains the human behavioural patterns, e.g. the so-called *Need and Greed principle* that states "Your needs and desires make you vulnerable. Once hustlers know what you really want, they can easily manipulate you."



3. Each player draws 3 cards from the set of *attack techniques*.

   The card deck contains attack techniques, e.g. the technique of reverse social engineering that comprises creating a problem for the selected person and solving it. The gained trust is used to ask the victim for a favour.



4. Each player gets 1 *attacker type* card.

   The card has two sides. One for an inside attacker that is a well known member of the organisation and has established trust. Another one for an outside attacker that is unknown to the members of the organisation and that has to establish trust.

Brainstorming Phase

5. The players take the role of the attacker. Each player thinks of how to apply the exploit of the behavioural pattern in combination with one of the three attacks on one of the persons in the overview diagram to attack an asset. Moreover, the player has to choose if she is an insider or outsider of the organisation.
   The players get **5 minutes** to think about their attacks.

Attack Phase

6. The active player presents his attack to the group. Each attack consists of a principle, an attack scenario, an attacker, a victim, a communication channel and a targeted asset. Note that after a player has proposed an attack it is finalised and cannot be changed anymore by the player.

Discussion

7. In this round the other players discuss if the proposed attack is feasible and bring arguments why this could be unrealistic. All attacks have to be documented. If the proposed attack is not plausible the turn ends immediately. Finally, the other players have to make the choice on how many points are granted. In addition, the other players can also propose improved versions of an attack and gain points.

Points

The following points can be gained per round:

Attack **2 P.** feasible | **1 P.** feasible with help
1 P. plausible but infeasible | **0 P.** non plausible → end turn
Attacker **2 P.** outside attacker | **1 P.** inside attacker
Principle **2 P.** perfect match | **1 P.** somewhat match | **0 P.** no match
Scenario **1 P.** match | **0 P.** no match
Attack Improvement by other Players
2 P. major improvement | **1 P.** minor improvement

Iterate (Phases 2 – 7)

8. The next player proposes an attack in the same fashion explained above. This iterates until all persons iterated at least twice. After each round, the players restock their cards. The Brainstorming Phase may be shortened by the players.

Debriefing

9. We propose the following steps for a structured threat elicitation:
   - ▷ Identify the most relevant targets of social engineers in your organisation
   - ▷ Try to figure out why some people were attacked more often and others not at all
   - ▷ Analyse why some communication channels were used more often than others
   - ▷ Determine which assets were attacked more often than others

Supported by:

Bayerisches Staatsministerium für Wirtschaft und Medien, Energie und Technologie

Federal Ministry of Education and Research

{kristian.beckers,veronika.fries}@tum.de

Bristish HCI Conference 2016

sebastian.pape@m-chair.de

A.3 A systematic Gap Analysis of Social Engineering Defence Mechanisms considering Social Psychology

Peter Schaab, Kristian Beckers, and Sebastian Pape. A systematic gap analysis of social engineering defence mechanisms considering social psychology. In *10th International Symposium on Human Aspects of Information Security & Assurance, HAISA 2016, Frankfurt, Germany, July 19-21, 2016, Proceedings.*, 2016. URL `https://www.cscan.org/openaccess/?paperid=301`

Proceedings of the Tenth International Symposium on
Human Aspects of Information Security & Assurance (HAISA 2016)

A systematic Gap Analysis of Social Engineering Defence Mechanisms Considering Social Psychology

P. Schaab[1], K. Beckers[1] and Sebastian Pape[2]

[1]Technische Universität München (TUM)
[2]Goethe Universität Frankfurt
e-mail: {peter.schaab, beckersk}@in.tum.de; Sebastian.Pape@m-chair.de

Abstract

Social engineering is the acquisition of information about computer systems by methods that deeply include non-technical means. While technical security of most critical systems is high, the systems remain vulnerable to attacks from social engineers. Social engineering is a technique that: (i) does not require any (advanced) technical tools, (ii) can be used by anyone, (iii) is cheap. Traditional penetration testing approaches often focus on vulnerabilities in network or software systems. Few approaches even consider the exploitation of humans via social engineering. While the amount of social engineering attacks and the damage they cause rise every year, the defences against social engineering do not evolve accordingly. Hence, the security awareness of these attacks by employees remains low. We examined the psychological principles of social engineering and which psychological techniques induce resistance to persuasion applicable for social engineering. The techniques examined are an enhancement of persuasion knowledge, attitude bolstering and influencing the decision making. While research exists elaborating on security awareness, the integration of resistance against persuasion has not been done. Therefore, we analysed current defence mechanisms and provide a gap analysis based on research in social psychology. Based on our findings we provide guidelines of how to improve social engineering defence mechanisms such as security awareness programs.

Keywords

social engineering, security management, persuasion, human-centred defence mechanisms

1. Introduction

Although security technology improves, the human user remains the weakest link in system security. Therefore, it is widely accepted that the people of an organization are the main vulnerability of any organization's security, as well as the most challenging aspect of system security (Mitnick and Simon, 2011). This is emphasized by many security consultants, as well as from genuine attackers, which accessed critical information via social engineering (Gragg, 2003). Early on Gulati (2003) reported that cyber attacks cost U.S. companies $266 million every year and that 80% of the attacks are a form of social engineering. A study in 2011 showed that nearly half of the considered large companies and a third of small companies fell victim of 25 or more social engineering attacks in the two years before (Dimensional Research, 2011). The study further shows that costs per incident usually vary

Proceedings of the Tenth International Symposium on
Human Aspects of Information Security & Assurance (HAISA 2016)

between $25 000 and over $100 000. Furthermore, surveys, like Verizon's 'Data Breach Investigation Report' (2012; 2013), show the impact of social engineering. Even though the awareness about the phenomenon of social engineering has increased, at least in literature, the impact has grown from 7% of breaches in 2012 to 29% of breaches in 2013 according to these studies. In addition, current security awareness programs are apparently ineffective (Pfleeger et al., 2014). These alarming numbers question whether the existing approaches towards awareness and defence of social engineering are fundamentally incomplete.

Frangopoulos et al. (2010) consider the psychological aspects of social engineering and relate them to persuasion techniques in their 2010 publication. In contrast to our work their work is not based on a literature review of behaviour psychology, but based on the expertise of the authors. Moreover, the scope of the authors is broader and consider physical measures, as well as security standards in their work. Our results classify existing research in IT security and persuasion in literature and contribute a structured gap analysis. In addition, Frangopoulos et al. (2012) transfer the knowledge of psychosocial risks, e.g. influence of headaches and colds on decisions, from a managerial and organisational point of view to the information security view.

Our hypothesis is that the psychological aspects behind social engineering and user psychology are not considered to their full extend. For instance, Ferreira et al. (2015) constitute psychological principles in social engineering and relate these principles to previous research of Cialdini (2009), Gragg (2003) and Stajano and Wilson (2011). However, these principles have to be the fundamental concern of any security defence mechanism against social engineering. Thus, we contribute a list of concepts that address social engineering defence mechanisms. We analyse in particular what IT security recommends in comparison to recommendations given by social psychology. The results of our analysis reveal fundamental gaps in today's security awareness approach. We provide a road map that shows how to address these gaps in the future. Our road map is an instrumental vision towards reducing the social engineering threat by addressing all relevant psychological aspects in its defence.

2. Methodology

Our research was guided by the methodology outlined in Fig. 1. We initialized the work with a working definition of social engineering (Sect. 3) and surveyed the state of the art from the viewpoint of computer science in particular with regard to IT security (Step 2) and separately from the viewpoint of social psychology (Sect. 4). We used the meta search engines Google Scholar and Scopus, which include the main libraries of IEEE, ACM, Springer, Elsevier and numerous further publishers. Based on the findings of our literature survey, we identified requirements and techniques from social sciences for defending against social engineering and map these to the defence mechanisms used in IT security today (Sect. 5). We outline the resulting gap and present a vision for overcoming these shortcomings of current IT security defences (Sect. 6). Finally, we conclude and provide directions for future research (Sect. 7).

Proceedings of the Tenth International Symposium on
Human Aspects of Information Security & Assurance (HAISA 2016)

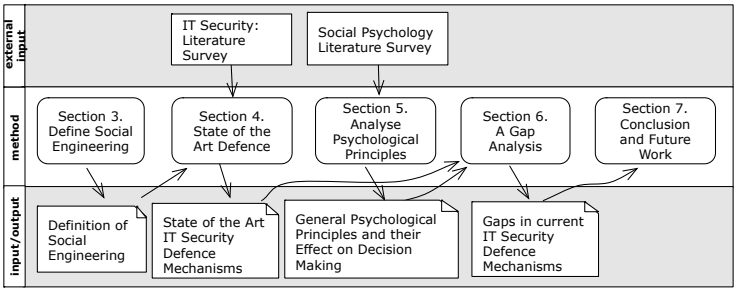


Figure 1: Methodology

3. Definition of Social Engineering

Although there is no agreed upon definition of social engineering, the common idea arising from the available definitions is that social engineering is the acquisition of confidential, private or privileged information by methods including both technical and non-technical means (Manske, 2009). This common idea is quite general, as it includes means of gaining information access such as shoulder surfing, dumpster diving, etc. However, it especially refers to social interaction as psychological process of manipulating or persuading people into disclosing such information (Thornburgh, 2004). Other than the former methods of accessing information, the latter are more complex and more difficult to resist, as persuasion is based on psychology. In this context, persuasion can be viewed as "any instance in which an active attempt is made to change a person's mind" (Petty and Cacioppo, 1996, p.4). The concept of 'optimism bias' states that people believe that others fall victim to misfortune, not themselves (Weinstein, 1980). Additionally, they tend to overestimate their possibilities to influence an event's outcome. Hence people think that they (i) will not be targeted by social engineering and (ii) are more likely to resist than their peers.

To actually raise resistance, we analyse how information security awareness can be increased. In alignment with Kruger and Kearny (2006) we define information security awareness as the degree to which employees understand the need for security measures and adjust their behaviour to prevent security incidents. Furthermore, in accordance with Veseli (2011) we focus on the information security dimensions attitude (how does a person feel about the topic) and behaviour (what does a person do) as they are an expression of conscious and unconscious knowledge (what does a person know).

4. An Analysis of Social Engineering Defence Mechanisms in IT Security

After having established the concept of social engineering, we analyse how the threat of social engineering is met in IT security. As the main vulnerability exploited by social engineering is inherent in human nature, it is the human element in systems that needs to be addressed. Thus, we concentrate on human based defence mechanisms. Predominantly three human based mitigation methods are proposed:

Proceedings of the Tenth International Symposium on
Human Aspects of Information Security & Assurance (HAISA 2016)

Policies, audits and security awareness programs, as indicated in Table 1. User awareness and security

Dimension		**Defence Mechanism**	**Description**
Knowledge	**Attitude**	Policy Compliance	- Foundation of information security - System standards and security levels - Guidelines for user behaviour
		Security Awareness Program	- Familiarity with security policy - Knowledge about sensitive, valuable information - Basic indicators, suspicious behaviour connected to social engineering attacks - (Recognition of being manipulated)
	Behaviour	Audit	- Test employee susceptibility to social engineering - Identify weaknesses of policy and security awareness program

Table 1: Defence mechanisms used in IT security

policies dominate the recommendations to defend social engineering (Scheeres, 2008).

Security Policies. Any information security is founded on its policy (Mitnick and Simon, 2011). Furthermore, policies provide instructions and guidelines how users should behave. It is especially hard to address social engineering in security policies, since people need to know how to respond to ambiguous requests (Gragg, 2003). By safe-guarding information, users should not come into uncertainty to decide whether certain information is sensitive or not. Necessarily these policies need to be enforced consistently throughout the system.

Security Awareness Programs. Upon establishment of a security policy all users need to be trained in security awareness programs to follow the policy, practices and procedures (Mitnick and Simon, 2011; Thornburgh, 2004). In general, the literature agrees upon the cornerstones of an awareness program. First of all, familiarity with the security policy needs to be established. It is important that everyone in the organization knows what kind of information is sensitive, hence particularly valuable for an attacker. Secondly, knowledge about social engineering is to be conveyed. This includes basics of social engineering, and how attacks work in detail. This should help employees to understand the reasons for related security policies that simply contains rules and usually not the reasoning behind it. The idea is that the understanding of why these polices were defined, will increase compliant behaviour among employees. In addition, the thought knowledge should reach beyond the rules in the policies and contain in particular indicators of social engineering attacks and what behaviour could be suspicious, such as requesting confidential information or to refuse provision of personal or contact information. Gragg (2003) demands the inclusion of additional training for key personnel to include inoculation, forewarning and reality check, see Section 5.

Audit. The conduction of audits is complementary to the above approaches (Thornburgh, 2004). It serves the purpose to test the susceptibility to social engineering attacks (Mitnick and Simon, 2011). Hence, it tests the effectiveness and identifies weaknesses of the other conducted methods (Winkler and Dealy, 1995). In this particular case, classic audits or penetration tests need an extension to social engineering penetration testing as done by Bakhshi et al. (2008). This extension is not trivial since it tests humans who can get upset and the work council needs to be involved.

5. Relevant Defence Mechanisms in Social Psychology

The intentions of security awareness programs are to inform about social engineering and sensitive information. It is assumed that by knowing about the threat of social engineering, users are less likely to be susceptible for such attacks. There is only a few researchers that have found this not to be sufficient, which appears to be ignored by most others. Gragg (2003) considers psychological principles of persuasion behind social engineering. Ferreira et al. (2015) have established a framework of psychological principles. These exhibit the ability to influence and potentially manipulate a person's attitude, believes and behaviour. Gragg therefore recommends techniques to build resistance against persuasion, borrowed from social psychology, to be included into awareness programs. An overview over these methods is given in

Dimension		**Defence Mechanism**	**Description**
Knowledge	**Attitude**	Persuasion Knowledge	- Information about tactics used in persuasion attempts and their potential influence on attitude and behaviour - Information about appropriate coping tactics
		Forewarning	- Warning of message content and persuasion attempt
		Attitude Bolstering	- Thought process strengthening security attitude
		Reality Check	- Demonstration of vulnerability to perceive risk of persuasion
	Behaviour	Inoculation	- Exposition to persuasive attempts and arguments of a social engineer - Provision of counter arguments to resist persuasion
		Decision Making	- Repeated exposition to "similar" decision making situations

Table 2: Defence mechanisms against persuasion borrowed from social psychology

Inoculation. A user gets exposed to persuasive attempts of a social engineer, he is put into a situation a social engineer would put him in. Thereby he is exposed to

Proceedings of the Tenth International Symposium on
Human Aspects of Information Security & Assurance (HAISA 2016)

arguments that a social engineer may use. Also he is given counter arguments that he can use to resist the persuasion. This works the same way as preventing a disease being spread by using inoculation and induces resistance to persuasion.

Forewarning. Forewarnings of message content and the persuasion attempt of the message triggers resistance to a social engineering attack. The intention is to not only warn about the persuasive attempt of a social engineer, but in particular to warn about the arguments being manipulative and deceptive. An example of this technique would be the warning about fraudulent IT support calls asking for user login and password.

Reality Check. As people tend to believe that they are invulnerable due to optimism bias, users need to realize that in fact they are vulnerable. Therefore, it has to be demonstrated to them, that they are vulnerable, to make them perceive the risks and training to be effective. However, any such effort has to be careful not to cause an amount of frustration that leads people to conclude their security efforts are useless. The balance between the demonstration of the vulnerability and the ensurance that people can make a difference in social engineering defence is vital for the success of defences.

Even though it appears that most programs are not extensive or limited in impact, it is unclear how much attention is given to these proposals in security practice. Nevertheless, research in the field of psychology over the past five decades has proven that inoculation is the most consistent and reliable method to induce resistance to persuasion (Miller et al., 2013). We are not aware of any study directly analysing the effects of inoculation to the resistance to social engineering. We are convinced that the principles behind inoculation are sound and we will analyse their effect on people in a future empirical study. In addition, Gragg (2003) has already adopted inoculation as a valuable mechanism for resistance to social engineering. Nevertheless, there exist further techniques in social psychology to train resistance to persuasion:

Persuasion Knowledge. Aim of security awareness programs is for users to experience resistance toward persuasion in case of a social engineering at- tack. This experience is increased if a user is concerned about being deceived (Friestad and Wright, 1994). Persuasion knowledge consists of information about tactics used in persuasive situations, their possible influence on attitudes and behaviour, their effectiveness and appropriateness, the persuasive agent's motives, and coping strategies (Fransen et al., 2015; Friestad & Wright, 1994). Activated persuasion knowledge usually either elicits suspicion about the persuasive agent's motives, or scepticism about arguments, and perceptions of manipulation or deception. Furthermore, it directs to options how to respond and selects coping tactics believed to be appropriate (Friestad and Wright, 1994). This positive relationship between persuasion knowledge and resistance to persuasive attempts is demonstrated by (Briñol et al., 2015): People are aware of persuasive attempts when having knowledge about persuasion and respond appropriately. This means educating users not only about common social engineering attack methods (e.g. phishing) but

particularly about psychological principles used in social engineering is an absolute necessity. As people also enhance their persuasion knowledge from experiences in social interactions, inoculation plays a vital role. Knowledge about coping tactics is, as indicated, essential to evaluate response options and to cope with persuasive attempts.

Attitude Bolstering. Awareness and knowledge of security policy, its implications and guidelines about e.g. confidential information are necessary to make use of attitude bolstering. The self or existing believes and attitudes are strengthened and therefore the vulnerability to persuasive attempts can be reduced (Fransen et al., 2015). In this process people generate thoughts that support their attitudes (Lydon et al., 1988). As demonstrated by Xu and Wyer (2012) it is possible to generate a bolstering mind-set that decreases the effectiveness of persuasive attempts. This is even possible when the cognitive behaviour leading to this bolstering mind-set has been performed in an unrelated, earlier situation.

Decision Making. Information is processed by using two different systems as explained by Kahneman (2003): intuition and reasoning. Decisions are made based on either one. Butavicius et al. (2015) found the preference for a decision making style has a link to the susceptibility to persuasion, i.e. phishing. Decisions based on heuristics or mental shortcuts are intuitive, impulsive judgements that are more likely to be influenced by persuasive attempts. But interestingly it seems that the style of decision making can be modified by training. This would imply that recurring exposure to different social engineering approaches helps in establishing effective strategies to cope with social engineering. Furthermore, it demonstrates that solely education about the threats of social engineering is not sufficient.

6. A Gap Analysis of Missing Defence Mechanisms in IT Security against Social Engineering

As indicated above, the available defence mechanisms can be classified into the dimensions attitude and behaviour, which in turn exert knowledge. Table 3 presents a mapping of defence mechanisms comparing suggestions in IT security against techniques known in social psychology. When comparing the dimension attitude, the limited scope of IT security becomes evident. As established in Section 4, in the dimension attitude IT security considers establishment of policy and security awareness programs. The purpose of security awareness programs is twofold. Firstly, it is concerned with getting users to know and adhere to the established policy. Secondly, security awareness program's scope is usually limited to the provision of basic knowledge about social engineering. In comparison social psychology offers distinctively more. Although some approaches may be at least partly covered. Forewarning can be seen as included in the education of social engineering basics, as malicious intention of social engineers certainly belongs to basic knowledge about social engineering. But persuasion knowledge goes beyond social engineering basics as it includes knowledge about persuasion strategies as well as counter tactics to rely on in any persuasive situation. For reliance on attitude bolstering good knowledge about security policy is necessary. Again IT security does the first step in user

education, but fails in the second step, the enhancement of this knowledge. The use of attitude bolstering, implies not only the knowledge about policy but its implications and a thought process initiated by each user that strengthens his attitude to e.g. keep sensitive information private. The necessity to perform a reality check can directly be deduced from the concept of 'optimism bias', as illustrated in Section 243. It might partially be covered in security awareness programs. A reality check might be done for e.g. spam mails. But as this particular reality check has a technical background and people tend to dismiss their possible failure by it being a technical detail and in the same time greatly underestimating personal susceptibility, it is important to demonstrate to them their failure in a non-technical environment as well.

Dimension		**IT Defence Mechanisms**	**Psychological Defence Mechanisms**
Knowledge	**Attitude**	Policy Compliance	-
		Security Awareness Program	Forewarning
		-	Persuasion Knowledge
		-	Attitude Bolstering
		-	Reality Check
	Behaviour	Audit	-
		-	Inoculation
		-	Decision Making

Table 3: Comparison of defence mechanisms suggested in IT security and social psychology

Table 3 presents another crucial finding. The dimension behaviour is under-represented in IT security. The only suggestion made for this dimension is to verify correct behaviour via audits. But IT security fails to actually enhance secure behaviour. Training correct behaviour as part of security awareness programs is, as indicated in Section 4, recommended by only a few authors and is usually at most done for spam mails. Even though this is the application of inoculation, this is only one possible social engineering attack and a particular technical one as well. Focus should again also be set on the persuasive nature of social engineering attacks. Hence trainings could for example include role plays. Additionally, it has been proven effective to alter the decision making process by conducting decision trainings where users make a "similar" decision in various appearances.

7. Conclusions and Future Work

Previously, we have discussed gaps in IT security. As indicated, both dimensions, attitude and behaviour, are represented inadequately in IT security when compared to recommendations from social psychology. To counter this gap. We envision a two-step improvement of available security awareness programs (as shown in Table 4). In a first step persuasion resistance trainings should be conducted. They should include a broad approach to social engineering including psychological principles and their effects, possible counter strategies, the initiation of attitude bolstering. As optimism bias is a strong enabler of successful social engineering, it would be desirable to

demonstrate users their susceptibility. This step is particularly promising, as it is feasible with little monetary effort. The second step is persuasive situation role plays. It is conceivable to include experiential exercises in this step as well as repeated decision trainings that force users to re-evaluate their knowledge and attitude by making a "similar" decision multiple times. This step is more effortful and it might suffice to only educate key personnel as it includes "live" training sessions guided by possibly costly trainers, actors or generally personnel capable of create persuasive situations.

Dimensions	Future defence mechanisms
Attitude	Persuasion resistance training
Behaviour	Persuasive situation role plays

Table 4: Envisioned training steps as part of security awareness programs

8. References

Bakhshi, T., Papadaki, M. and Furnell, S., 2008. A Practical Assessment of Social Engineering Vulnerabilities. In N. L. Clarke & S. Furnell, eds. *2nd International Conference on Human Aspects of Information Security and Assurance, {HAISA} 2008, Plymouth, UK, July 8-9, 2008. Proceedings*. University of Plymouth, pp. 12–23.

Briñol, P., Rucker, D.D. and Petty, R.E., 2015. Naïve theories about persuasion: Implications for information processing and consumer attitude change. *International Journal of Advertising*, 34(1), pp.85–106.

Butavicius, M. et al., 2015. Breaching the Human Firewall : Social engineering in Phishing and Spear-Phishing Emails. *Australasian Conference on Information Systems*, pp.1–11.

Cialdini, R.B., 2009. *Influence: the psychology of persuasion* EPub editi., New York: Collins.

Dimensional Research, 2011. *The Risk of Social Engineering on Information Security: A Survey of IT Professionals*, 2011

Ferreira, A., Coventry, L. and Lenzini, G., 2015. Principles of Persuasion in Social Engineering and Their Use in Phishing. In T. Tryfonas & I. Askoxylakis, eds. *Human Aspects of Information Security, Privacy, and Trust SE - 4*. Lecture Notes in Computer Science. Springer International Publishing, pp. 36–47. Available at:

Frangopoulos E.D.; Eloff, M.M.; Venter L.M., 2010. Psychological considerations in Social Engineering - The "ψ-wall" as defense, Proceedings of the IADIS International Conference Information Systems, pp. 1-20.

Frangopoulos, E.D., Eloff, M.M. and Venter, L.M., 2012. Psychosocial Risks: can their effects on the Security of Information Systems really be ignored? In N. L. Clarke & S. Furnell, eds. *6th International Symposium on Human Aspects of Information Security and Assurance, {HAISA} 2012, Crete, Greece, June 6-8, 2012. Proceedings*. University of Plymouth, pp. 52–63.

Fransen, M.L. et al., 2015. Strategies and motives for resistance to persuasion : an integrative framework. *Frontiers in psychology*, 6(August), pp.1–12.

Friestad, M. and Wright, P., 1994. The Persuasion Knowledge Model: How People Cope with Persuasion Attempts. *Journal of Consumer Research*, 21(1), pp.1–31.

Gragg, D., 2003. A multi-level defense against social engineering. *SANS Reading Room*.

Gulati, R., 2003. The Threat of Social Engineering and your defense against it. *SANS Reading Room*.

Kahneman, D., 2003. A perspective on judgment and choice: mapping bounded rationality. *The American psychologist*, 58(9), pp.697–720.

Kruger, H. A. and Kearney, W. D., 2006. A prototype for assessing information security awareness. Comput. Secur. 25, 4 , pp. 289-296.

Lydon, J., Zanna, M.P. and Ross, M., 1988. Bolstering Attitudes by Autobiographical Recall: Attitude Persistence and Selective Memory. *Personality and Social Psychology Bulletin*, 14(1), pp.78–86. Available at: http://psp.sagepub.com/content/14/1/78.abstract.

Manske, K., 2009. An Introduction to Social Engineering. *Information Security Journal: A Global Perspective*, 9(5), pp.1–7.

Miller, C.H. et al., 2013. Boosting the Potency of Resistance: Combining the Motivational Forces of Inoculation and Psychological Reactance. *Human Communication Research*, 39(1), pp.127–155.

Mitnick, K.D. and Simon, W.L., 2011. *The art of deception: Controlling the human element of security*, John Wiley & Sons.

Petty, R.E. and Cacioppo, J.T., 1996. *Attitudes and persuasion: Classic and contemporary approaches*, Boulder, CO, US: Westview Press.

Pfleeger, S.L., Sasse, M.A. and Furnham, A., 2014. From Weakest Link to Security Hero: Transforming Staff Security Behavior. *Journal of Homeland Security and Emergency Management*, 11(4), pp.489–510.

Sagarin, B.J. et al., 2002. Dispelling the illusion of invulnerability: The motivations and mechanisms of resistance to persuasion. *Journal of Personality and Social Psychology*, 83(3), pp.526–541.

Scheeres, J.W., 2008. *Establishing the human firewall: reducing an individual's vulnerability to social engineering attacks*,

Stajano, F. and Wilson, P., 2011. Understanding Scam Victims: Seven Principles for Systems Security. *Commun. ACM*, 54(3), pp.70–75.

Thornburgh, T., 2004. Social Engineering: The "Dark Art." In *Proceedings of the 1st Annual Conference on Information Security Curriculum Development*. InfoSecCD '04. New York, NY, USA: ACM, pp. 133–135.

Verizon, 2012. Data Breach Investigations Report. Available at: http://www.verizonenterprise.com/resources/reports/rp_data-breach-investigations-report-2012-ebk_en_xg.pdf [Accessed January 13, 2016].

Verizon, 2013. Data Breach Investigations Report. Available at: http://www.verizonenterprise.com/resources/reports/rp_data-breach-investigations-report-2013_en_xg.pdf [Accessed January 13, 2016].

Veseli, I., 2011. *Measuring the Effectiveness of Information Security Awareness Program*. Gjøvik University College.

Weinstein, N.D., 1980. Unrealistic Optimism About Future Life events. *Journal of Personality and Social Psychology*, 39(5), pp.806–820.

Winkler, I.S. and Dealy, B., 1995. Information Security Technology?...Don't Rely on It A Case Study in Social Engineering. In *Fifth Usenix Security Symposium*. pp. 1–6.

Xu, A.J. and Wyer, R.S.J., 2012. The Role of Bolstering and Counterarguing Mind-Sets in Persuasion. *Journal of Consumer Research*, 38(5), pp.920–932. Available at: http://www.jstor.org/stable/10.1086/661112.

A.4 Social engineering defence mechanisms and counteracting training strategies

Peter Schaab, Kristian Beckers, and Sebastian Pape. Social engineering defence mechanisms and counteracting training strategies. *Information and Computer Security*, 25(2):206–222, 2017. doi: 10.1108/ICS-04-2017-0022. URL https://doi.org/10.1108/ICS-04-2017-0022

Social Engineering Defence Mechanisms and counteracting Training Strategies

Peter Schaab[*], Kristian Beckers[*], Sebastian Pape[+]

[*]Technische Universität München (TUM)
peter.schaab@in.tum.de, beckersk@in.tum.de
[+]Goethe Universität Frankfurt
Sebastian.Pape@m-chair.de

Abstract

Purpose – The paper aims to outline strategies for defence against social engineering that are missing in current best practices of IT security. Reason for the incomplete training techniques in IT security is the interdisciplinary of the field. Social engineering is focusing on exploiting human behaviour and this is not sufficiently addressed in IT security. Instead most defence strategies are devised by IT security experts with a background in information systems rather than human behaviour. We aim to outline this gap and point out strategies to fill the gaps.
Design/methodology/approach – We conducted a literature review from viewpoint IT security and viewpoint social psychology. In addition, we mapped the results to outline gaps and analysed how these gaps could be filled using established methods from social psychology and discussed our findings.
Findings – We analysed gaps in social engineering defences and mapped them to underlying psychological principles of social engineering attacks e.g. social proof. Furthermore, we discuss which type of countermeasure proposed in social psychology should be applied to counteract which principle. We derived two training strategies from these results that go beyond the state of the art trainings in IT security and allow security professional to raise companies' bars against social engineering attacks.
Originality/value – Our training strategies outline how interdisciplinary research between computer science and social psychology can lead to a more complete defence against social engineering by providing reference points for researchers and IT security professional with advice on how to improve training.

Keywords

social engineering, security management, persuasion, human-centred defence mechanisms

Paper type

Literature review

1. Introduction

Although security technology improves, the human user remains the weakest link in system security. Therefore, it is widely accepted that the people of an organization are the main vulnerability of any organization's security, as well as the most challenging aspect of system security (Barrett, 2003; Mitnick and Simon, 2011). This is emphasized by many security consultants, as well as from genuine attackers, which accessed critical information via social engineering (Gragg, 2003; Warkentin and Willison, 2009). Early on Gulati (2003) reported that cyber attacks cost U.S. companies $266 million every year and that 80% of the attacks are a form of social engineering. A study in 2011 showed that nearly half of the considered large companies and a third of small companies fell victim of 25 or more social engineering attacks in the two years before (Dimensional Research, 2011). The study further shows that costs per incident usually vary between $25 000 and over $100 000. Furthermore, surveys, like Verizon's 'Data Breach Investigation Report' (2012; 2013), show the impact of social engineering. Even though the awareness about the phenomenon of social engineering has increased, at least in literature, the impact has grown from 7% of breaches in 2012 to 29% of breaches in 2013 according to these studies. In addition, current security awareness programs are apparently ineffective (Pfleeger et al., 2014). These alarming numbers question whether the existing approaches towards awareness and defence of social engineering are fundamentally incomplete.

Frangopoulos et al. (2010) consider the psychological aspects of social engineering and relate them to persuasion techniques in their 2010 publication. In contrast to our work their work is not based on a literature review of behaviour psychology, but based on the expertise of the authors. Moreover, the scope of the authors is broader and consider physical measures, as well as security standards in their work. Our results classify existing research in IT security and persuasion in literature and contribute a structured gap analysis. In addition, Frangopoulos et al. (2012) transfer the knowledge of psychosocial risks, e.g. influence of headaches and colds on decisions, from a managerial and organisational point of view to the information security view.

Our hypothesis is that the psychological aspects behind social engineering and user psychology are not considered to their full extend. For instance, Ferreira et al. (2015) constitute psychological principles in social engineering and relate these principles to previous research of Cialdini (2009), Gragg (2003) and Stajano and Wilson (2011). Thus, as starting point we analysed the psychological explanations of these social engineering principles by relating the insights of Cialdini, Gragg and Stajano and Wilson. As these principles have to be the fundamental concern of any security defence mechanism against social engineering, we contribute a list of concepts that address social engineering defence mechanisms. In particular, we analyse recommendations from IT in comparison to recommendations given by social psychology. The results of our analysis are twofold. On one side we provide a mapping between the influence of the identified defence mechanism to mitigate social engineering attacks based on the individual psychological principles. On the other side the analysis reveals fundamental gaps in today's security awareness approach. We provide a road map that shows how to address these gaps in the future. Our road map

is an instrumental vision towards reducing the social engineering threat by addressing all relevant psychological aspects in its defence.

2. Methodology

Our research was guided by the methodology outlined in Fig. 1. We initialized the work with a working definition of social engineering (Sect. 3) and surveyed the state of the art from the viewpoint of computer science in particular with regard to IT security (Sect. 4) and separately from the viewpoint of social psychology (Sect. 6). We used the meta search engines Google Scholar and Scopus, which include the main libraries of IEEE, ACM, Springer, Elsevier and numerous further publishers. Based on the findings of our literature survey and a review of psychological principles behind social engineering (Sect. 5), we identified requirements and techniques from social sciences for defending against social engineering (Sect. 6) and map these to underlying psychological principles of the attacks (Sect. 7). Next, we map these to the defence mechanisms used in IT security today (Sect. 8). We outline the resulting gap and present a vision for overcoming these shortcomings of current IT security defences and derive missing training strategies (Sect. 9). Finally, we discuss our results and provide directions for future research.

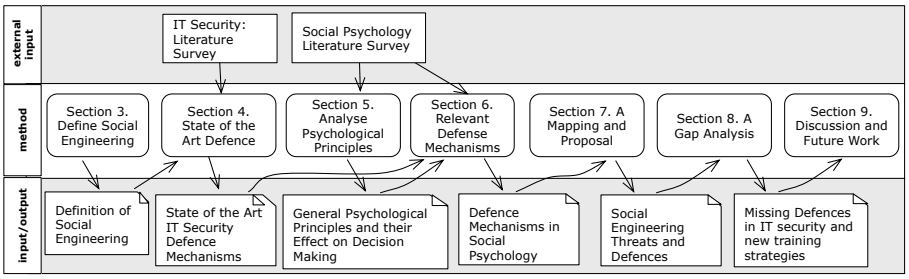


Figure 1: Methodology

3. Definition of Social Engineering

Although there is no agreed upon definition of social engineering, the common idea arising from the available definitions is that social engineering is the acquisition of confidential, private or privileged information by methods including both technical and non-technical means (Manske, 2009). This common idea is quite general, as it includes means of gaining information access such as shoulder surfing, dumpster diving, etc. However, it especially refers to social interaction as psychological process of manipulating or persuading people into disclosing such information (Thornburgh, 2004). Other than the former methods of accessing information, the latter are more complex and more difficult to resist, as persuasion is based on psychology. In this context, persuasion can be viewed as "any instance in which an active attempt is made to change a person's mind" (Petty and Cacioppo, 1996, p.4). The concept of 'optimism

bias' states that people believe that others fall victim to misfortune, not themselves (Weinstein, 1980). Additionally, they tend to overestimate their possibilities to influence an event's outcome. Hence people think that they (i) will not be targeted by social engineering and (ii) are more likely to resist than their peers.

To actually raise resistance, we analyse how information security awareness can be increased. In alignment with Kruger and Kearny (2006) we define information security awareness as the degree to which employees understand the need for security measures and adjust their behaviour to prevent security incidents. Furthermore, in accordance with Veseli (2011) we focus on the information security dimensions attitude (how does a person feel about the topic) and behaviour (what does a person do) as they are an expression of conscious and unconscious knowledge (what does a person know).

4. Analysis of Social Engineering Defence Mechanisms in IT Security

After having established the concept of social engineering, we analyse how the threat of social engineering is met in IT security. As the main vulnerability exploited by social engineering is inherent in human nature, it is the human element in systems that needs to be addressed. Thus, we concentrate on human based defence mechanisms. Predominantly three human based mitigation methods are proposed: Policies, audits and security awareness programs, as indicated in Table 1. User awareness and security policies dominate the recommendations to defend social engineering (Scheeres, 2008).

Dimension		**Defence Mechanism**	**Description**
Knowledge	**Attitude**	Policy Compliance	- Foundation of information security - System standards and security levels - Guidelines for user behaviour
		Security Awareness Program	- Familiarity with security policy - Knowledge about sensitive, valuable information - Basic indicators, suspicious behaviour connected to social engineering attacks - (Recognition of being manipulated)
	Behaviour	Audit	- Test employee susceptibility to social engineering - Identify weaknesses of policy and security

			awareness program

Table 1: Defence mechanisms used in IT security

Security Policies. Any information security is founded on its policy (Mitnick and Simon, 2011). Furthermore, policies provide instructions and guidelines how users should behave. It is especially hard to address social engineering in security policies, since people need to know how to respond to ambiguous requests (Gragg, 2003). By safe-guarding information, users should not come into uncertainty to decide whether certain information is sensitive or not. Necessarily these policies need to be enforced consistently throughout the system.

Security Awareness Programs. Upon establishment of a security policy all users need to be trained in security awareness programs to follow the policy, practices and procedures (Mitnick and Simon, 2011; Thornburgh, 2004). In general, the literature agrees upon the cornerstones of an awareness program. First of all, familiarity with the security policy needs to be established. It is important that everyone in the organization knows what kind of information is sensitive, hence particularly valuable for an attacker. Secondly, knowledge about social engineering is to be conveyed. This includes basics of social engineering, and how attacks work in detail. This should help employees to understand the reasons for related security policies that simply contains rules and usually not the reasoning behind it. The idea is that the understanding of why these polices were defined, will increase compliant behaviour among employees. In addition, the thought knowledge should reach beyond the rules in the policies and contain in particular indicators of social engineering attacks and what behaviour could be suspicious, such as requesting confidential information or to refuse provision of personal or contact information. Gragg (2003) demands the inclusion of additional training for key personnel to include inoculation, forewarning and reality check, see Section 6.

Audit. The conduction of audits is complementary to the above approaches (Thornburgh, 2004). It serves the purpose to test the susceptibility to social engineering attacks (Mitnick and Simon, 2011). Hence, it tests the effectiveness and identifies weaknesses of the other conducted methods (Winkler and Dealy, 1995). In this particular case, classic audits or penetration tests need an extension to social engineering penetration testing as done by Bakhshi et al. (2008). This extension is not trivial since it tests humans who can get upset and the work council needs to be involved.

5. Analysis of Psychological Principles underlying Social Engineering

According to Rusch (1999) two ways of persuading an individual exist:

1. A central route to persuasion based on sound analytical reasoning of facts;

2. or a peripheral route to persuasion relying on acceptance without deeply reasoning about the facts by triggering mental shortcuts or eliciting emotions.[1]

Thornburgh (2004) and Ivaturi and Janczewski (2011) state that the central route (1) is not a real option for a social engineer as his entire approach is based on "misrepresentation and dissembling" (Thornburgh, 2004) meaning the employment of deception and manipulation. Gragg (2003), analysed literature on persuasion, influence and social engineering and suggests seven psychological triggers which are explicitly referred to as being applicable to social engineering. Scheeres (2008) has deduced that Gragg's triggers are in line with the principles of Cialdini (2009). This means Cialdini's result is also applicable to social engineering. Additionally, it is generally accepted that the same psychological techniques are applied in social engineering as in traditional fraud (Rusch, 1999). Therefore, Stajano and Wilson (2011) identified seven principles of scam applicable to social engineering. Based on these findings Ferreira et al. (2015) enhance the principles of Cialdini, Gragg, and Stajano and Wilson by constituting a complete set of psychological principles of persuasion in social engineering. Ferreira et al. related the already existing principles to each other and identified five principles of persuasion in social engineering that account for all available principles. We investigate why people are prone for them to get insight into the prevalent threats of social engineering. We do not analyse studies that suggest or validate the behaviours described.[2] Instead we focus on the triggered behaviours and try to find insights into their functioning. This is done to gain further valuable understanding of the triggers to find valid countermeasures in a next step. A summary of these principles is provided in Table 2.

Psychological Principle	**Description**
Authority	- Conditioning to respond to authority - Beneficial to unconditionally conform to authority - Authority indicated by abstract rank
Social Proof	- Reliance on majority's apparent behaviour in determining appropriate behaviour in uncertainty - Confidence when seemingly not solely responsible

[1] These two routes are referred to as System 1 and System 2 in cognitive psychology.

[2] Valid examples, studies and a variety of scenarios, which principle is applicable when, can be found aplenty in Cialdini (2009), Gragg (2003) and Stajano and Wilson (2011).

Liking, Similarity and Deception	- Tendency to react positively to whom some kind of 'relationship' has been established - Relationship sources: attractiveness, compliments, familiarity, liking - Satisfaction of expectations through manipulation
Commitment, Reciprocation and Consistency	- Urge to consistency with commitment - Societal obligation to future repayments of received concessions
Distraction	- Limited attention is focused on seemingly important facts or actions - Directing attention in desired direction by manipulation of focus

Table 2: Psychological principles of social engineering

Authority. "Society trains people not to question authority so they are conditioned to respond to it" (Cialdini, 2009).As Milgram (1974) puts it, conforming to authority figures' wishes and commands has always proved to be beneficial for us. As long as we can think these people (e.g. parents, teachers) knew more than us, and for us taking advice had advantages — partly due to greater wisdom, partly due to the control of rewards and punishments (Milgram, 1974). This pattern persists up to adulthood, only authority figures change, now appearing as e.g. employers or judges. But it continuously might be wise to comply with the dictate of constituted authorities, independently of how this authority constitutes itself. In modern society responses to authority are made to abstract rank, even in the absence of any substance of authority, as long as it is indicated by an insignia, uniform or title (Cialdini, 2009; Milgram, 1974). Due to this societal trained behaviour of unconditioned response to authority, people without questioning adhere to the dictate of authoritative figures as demonstrated by the famous Milgram (Milgram, 1963) experiment.[3] "People usually follow an expert or pretence of authority and do a great deal for someone they think represents authority" (Cialdini, 2009).

Social Proof. People rely on others in determining what is appropriate in any given

[3] In an experiment individuals were instructed to supervise electric shocks of increasing strength to other individuals when those made mistakes. The victims were accomplices who did not in fact receive the shocks. The individuals complied with shocking extent. They continued to apply electrical shocks of up to 450 V. Even when victims pretended screaming and fainting they did not spare the experimental subjects.

situation. According to Cialdini (2009) experience tells us to act according to social evidence rather than to its contrary. Especially when in uncertainty of correct behaviour, the behaviour of the majority of people tends to be correct and therefore constitutes correct behaviour for ourselves or at least provides a feeling of confidence and safety to conduct an otherwise doubtable action or an action against our self-interest (Cialdini, 2009; Rusch, 1999; Stajano and Wilson, 2011). Furthermore, the behaviour of people similar to us, more powerfully establishes what is considered correct. As pointed out by Stajano and Wilson (2011) and Gragg (2003) this principle also accounts for people's will to take risks in an action, especially if not being held solely responsible. "People let their guard and suspicion down when everyone else appears to share the same behaviours and risks. In this way, they will not be held solely responsible for their actions." (Cialdini, 2009)

Liking, Similarity and Deception. Humans have a tendency to abide and comply or at least react positively to whom some kind of 'relationship' exists or is established. This relationship can take a variety of manifestations. Cialdini (2009) describes the major mechanisms of deceiving an individual into one of these relationships:

Attractiveness. Physical attractiveness is a characteristic that is associated with a 'halo effect'.[4] And therefore people assign favourable traits such as kindness, honesty and trustworthiness to attractive persons and therefore treat these persons favourable.

Similarity. To have identical or similar characteristics with an individual incentivizes people to favour this individual. This similarity can be accomplished in a wide range of attributes, such as opinions, personality traits, personal interests, background, appearance, etc.

Compliments. People tend to react positively to praise, affinity or general compliments to such an extent as for liking and compliance.

Contact and Cooperation. Attitude, especially the favour, towards an individual is influenced by the exposition to it. Therefore, familiarity evoked by contact usually leads to a more favourable mindset. This can even be increased through mutual cooperation or the attempt to establish a 'we' or 'us' as Gragg (2003) points out as well.

Conditioning and Association. Simple association with bad or good things influences how people feel about someone, it is enough to stimulate either like or dislike (Lott

[4] A halo effect occurs when one characteristic of an individual dominates how this individual is perceived by others.

and Lott, 1965).

Besides deceiving an individual into one of the above relations, Stajano and Wilson (2011) indicate that by knowing people's expectations, an individual can be deceived into authenticating a person and therefore it can be manipulated into moving along within any situation as long as the individual's expectations are satisfied.

Commitment, Reciprocation and Consistency. People feel induced to be consistent once having committed (publically) to a specific action. This tendency is neither influenced by the commitment not being very wise, nor by recognizing it to be foolish or in contrast to our own interests (Cacioppo et al., 1986; Cialdini, 2009). As Stajano and Wilson (2011) and Gragg (2003) emphasize this also accounts for requests that may not have been legitimated or are even illegal. According to Cialdini (2009) people encounter personal and interpersonal pressures to stay consistent with an earlier commitment causing them to act accordingly to their previous commitment. People tend to take considerable pains to stay consistent (Rusch, 1999). Staying consistent is in fact considered as central motivator for human's behaviour as it is highly rewarded in our culture. It is associated with integrity, personal and intellectual strength, whereas inconsistency is viewed as untrustworthy and therefore an undesired personal characteristic. Consistency provides reasonable orientation to our lives. This is accompanied by the tendency to believe that others express their true feelings and attitudes when making a statement (Gragg, 2003).

The desire to appear consistent in our actions has formed another strongly connected behaviour or well-established rule in social interactions — reciprocation. This rule obligates an individual to future repayment for favours or generally any- thing given or promised to us (Cialdini, 2009; Rusch, 1999). According to Gouldner (1960) this rule is ingrained into any human society. As Cialdini (2009) puts it, a society wide shared feeling of future obligation is necessary to make social interaction in today's form possible, as it lowers natural inhibitions against transactions and instead allows an individual to provide resources with confidence that the given is not being lost but returned in the future. As this brings immense advantages, people are trained to comply and not question the rule of reciprocation. Again, society considers individuals that take and do not return anything with negativity and therefore it is inherent in human's desire to try and avoid this.

The above comprises certain implications, that distinguish the rule of reciprocation from the other principles (Cialdini, 2009):

– By imposing a favour on us a disliked or unwelcomed person enhances his chance of our compliance significantly.

– An uninvited favour causes a feeling of indebtedness, as receiving the favour obligates to repay. This enables others to choose who is indebted to them, not oneself.

– Although generally the rule encourages equal exchanges, it enables an individual

to choose both the kind of initial indebting favour, e.g. a small one, as well as the kind of compensating return favour, e.g. a significantly larger one.

– Furthermore, the rule implies the obligation of a concession, if someone has made an initial concession. Mutual concession promotes compromise in social interactions, as requirements of interacting persons often are unacceptable to one another.

Distraction People focus their limited attention on what is perceived to be most interesting or most important for a variety of reasons, and ignore seemingly uninteresting and unimportant facts or actions that may happen simultaneously (Stajano and Wilson, 2011). Due to this limited attention, it is possible to direct an individual in any desired direction, the individual is distracted. Basically these distractions heighten people's emotional state, which interferes with their ability to evaluate facts or actions by logical reasoning (Ferreira et al., 2015; Gragg, 2003). This can be achieved in a number of ways:

Human's Needs. Knowing a person's needs, desires and fears provides an understanding what drives him and how he behaves. This makes him vulnerable to emotional manipulation (Gragg, 2003; Stajano and Wilson, 2011). The phenomena is called counterfactual thinking and describes how the anticipation of future possibilities, caused by aiming at a person's needs, impedes reasoning (Landman and Petty, 2000).

Time. Depending on the urgency of a request the caused response may be different as it hinders evaluation (Stajano and Wilson, 2011). The same accounts for an information or sensory overload (Gragg, 2003). This is due to time not being available to process all information or implications of a request.

Scarcity. Potential loss highly influences decision making. By considering the availability of something people may often come to a decision about quality or worthiness without actually reasoning about e.g. their need (Lynn, 1989). Additionally, humans have a need to retain their freedom, thus in case a choice is limited or threatened the desire to preserve their freedom decidedly raises, as personal control is reduced (Brehm, 1966).

When people's attention is focused, directed or influenced by any of the above factors, they are distracted from proper evaluation and protection of their true intentions (Stajano and Wilson, 2011).

The analysed psychological principles share one special characteristic. They all describe how an individual or humans in general are induced to use a specific, automated decision mechanism, often called heuristic or mental shortcut, rather than rational reasoning. This is achieved by making use of the described concepts and human tendencies. After having analysed these tendencies and triggers, it is necessary to understand the different mechanisms in decision making. As Kahneman (2003) explains, humans cognitive functioning is distinguished into two separate cognitive systems. One system intuits (System 1) and the other reasons (System 2):

> "The operations of System 1 are typically fast, automatic, effortless, associative, implicit (not available to introspection), and often emotionally charged; they are also governed by habit and therefore difficult to control or modify. The operations of System 2 are slower, serial, effortful, more likely to be consciously monitored and deliberately controlled; they are also relatively flexible and potentially rule governed." (Kahneman, 2003)

Kahneman (2003) furthermore describes the differentiating aspects of the two systems. System 1 generates impressions of perceptions and thought, which are involuntarily and not necessarily verbally explicit. In comparison, judgments are intentional and explicit even when not verbally expressed. This means, when judging System 2 is usually involved, whether the judgment originates from impressions or reasoning. If a judgment directly reflects impressions and was not modified by System 2 then it is an intuitive judgment. Normally many intuitive judgements are expressed, even though System 2 is set to monitor mental operations (Gilbert, 2002; Stanovich and West, 2002). The competing behaviour of the two systems is summarized in Figure 2. As self-monitoring as well as reasoning are effortful operations, System 2 is affected by dual-task interference (Kahneman, 2003). Due to operations of System 2 being effortful, the monitoring of intuitive judgments usually is not very strict and therefore erroneous ones are not hindered because plausible judgments that are readily made are trusted (Cialdini, 2009; Kahneman, 2003). Being lax in monitoring is not only laziness or the attempt to avoid hard thinking, it is also a mechanism to reduce cognitive load. Besides by leaving System 1 in autopilot and not thinking straight troubling realizations can and will be avoided (Kahneman, 2003).

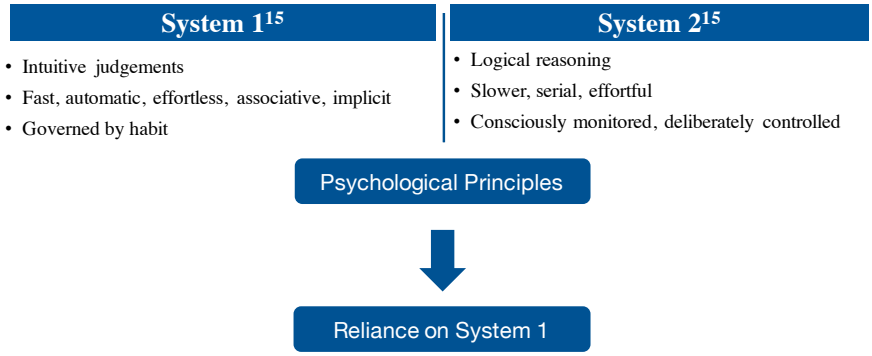


Figure 2: Psychological Triggers

Figure 2 illustrates that by using any of the psychological triggers, a social engineer tries to push the person opposite to rely on System 1, as there exists an evolutionary built heuristic that delivers an intuitive judgment, which is usually not monitored by

System 2.[5]

6. Relevant Defence Mechanisms in Social Psychology

The intentions of security awareness programs are to inform about social engineering and sensitive information. It is assumed that by knowing about the threat of social engineering, users are less likely to be susceptible for such attacks. There is only a few researchers that have found this not to be sufficient, which appears to be ignored by most others. Gragg (2003) considers psychological principles of persuasion behind social engineering. Ferreira et al. (2015) have established a framework of psychological principles. These exhibit the ability to influence and potentially manipulate a person's attitude, believes and behaviour. Gragg therefore recommends techniques to build resistance against persuasion, borrowed from social psychology, to be included into awareness programs. An overview over these methods is given in Table 3. They build on Sagarin et al. (2002):

Dimension		**Defence Mechanism**	**Description**
Knowledge	**Attitude**	Persuasion Knowledge	- Information about tactics used in persuasion attempts and their potential influence on attitude and behaviour - Information about appropriate coping
		Forewarning	- Warning of message content and persuasion attempt
		Attitude Bolstering	- Thought process strengthening security attitude
		Reality Check	- Demonstration of vulnerability to perceive risk of persuasion
	Behaviour	Inoculation	- Exposition to persuasive attempts and arguments of a social engineer - Provision of counter arguments to resist persuasion

[5] Of course there have evolved many more than the above introduced heuristics, allowing people to function effectively but therefore allowing people to bypass System 2 (Gigerenzer and Todd, 1999). We kept to the ones which are directly linked to persuading an individual. Although some of the available heuristics may have further impact on the behavior when attacked by a social engineer. For a collection of these heuristics view Schneier (2008) and Kahneman (2003).

		Decision Making	- Repeated exposition to "similar" decision making situations

Table 3: Defence mechanisms against persuasion borrowed from social psychology

Inoculation. A user gets exposed to persuasive attempts of a social engineer, he is put into a situation a social engineer would put him in. Thereby he is exposed to arguments that a social engineer may use. Also he is given counter arguments that he can use to resist the persuasion. This works the same way as preventing a disease being spread by using inoculation and induces resistance to persuasion.

Forewarning. Forewarnings of message content and the persuasion attempt of the message triggers resistance to a social engineering attack. The intention is to not only warn about the persuasive attempt of a social engineer, but in particular to warn about the arguments being manipulative and deceptive. An example of this technique would be the warning about fraudulent IT support calls asking for user login and password.

Reality Check. As people tend to believe that they are invulnerable due to optimism bias, users need to realize that in fact they are vulnerable. Therefore, it has to be demonstrated to them, that they are vulnerable, to make them perceive the risks and training to be effective. However, any such effort has to be careful not to cause an amount of frustration that leads people to conclude their security efforts are useless. The balance between the demonstration of the vulnerability and the ensurance that people can make a difference in social engineering defence is vital for the success of defences.

Even though it appears that most programs are not extensive or limited in impact, it is unclear how much attention is given to these proposals in security practice. Nevertheless, research in the field of psychology over the past five decades has proven that inoculation is the most consistent and reliable method to induce resistance to persuasion (Miller et al., 2013). We are not aware of any study directly analysing the effects of inoculation to the resistance to social engineering. We are convinced that the principles behind inoculation are sound and we will analyse their effect on people in a future empirical study. In addition, Gragg (2003) has already adopted inoculation as a valuable mechanism for resistance to social engineering. Nevertheless, there exist further techniques in social psychology to train resistance to persuasion:

Persuasion Knowledge. Aim of security awareness programs is for users to experience resistance toward persuasion in case of a social engineering at- tack. This experience is increased if a user is concerned about being deceived (Friestad and Wright, 1994). Persuasion knowledge consists of information about tactics used in persuasive situations, their possible influence on attitudes and behaviour, their effectiveness and appropriateness, the persuasive agent's motives, and coping strategies (Fransen et al., 2015; Friestad & Wright, 1994). Activated persuasion knowledge usually either elicits suspicion about the persuasive agent's motives, or scepticism about arguments, and perceptions of manipulation or deception. Furthermore, it directs to options how to respond and selects coping tactics believed to be appropriate (Friestad and Wright, 1994). This positive relationship between persuasion knowledge and resistance to persuasive attempts is demonstrated by

(Briñol et al., 2015): People are aware of persuasive attempts when having knowledge about persuasion and respond appropriately. This means educating users not only about common social engineering attack methods (e.g. phishing) but particularly about psychological principles used in social engineering is an absolute necessity. As people also enhance their persuasion knowledge from experiences in social interactions, inoculation plays a vital role. Knowledge about coping tactics is, as indicated, essential to evaluate response options and to cope with persuasive attempts.

Attitude Bolstering. Awareness and knowledge of security policy, its implications and guidelines about e.g. confidential information are necessary to make use of attitude bolstering. The self or existing believes and attitudes are strengthened and therefore the vulnerability to persuasive attempts can be reduced (Fransen et al., 2015). In this process people generate thoughts that support their attitudes (Lydon et al., 1988). As demonstrated by Xu and Wyer (2012) it is possible to generate a bolstering mind-set that decreases the effectiveness of persuasive attempts. This is even possible when the cognitive behaviour leading to this bolstering mind-set has been performed in an unrelated, earlier situation.

Decision Making. Information is processed by using two different systems as explained by Kahneman (2003): intuition and reasoning. Decisions are made based on either one. Butavicius et al. (2015) found the preference for a decision making style has a link to the susceptibility to persuasion, i.e. phishing. Decisions based on heuristics or mental shortcuts are intuitive, impulsive judgements that are more likely to be influenced by persuasive attempts. But interestingly it seems that the style of decision making can be modified by training. This would imply that recurring exposure to different social engineering approaches helps in establishing effective strategies to cope with social engineering. Furthermore, it demonstrates that solely education about the threats of social engineering is not sufficient.

7. Mapping of Defence Mechanisms against Psychological Principles

In order to get a better understanding how defence mechanisms work, we mapped them against the psychological principles (see Table 4).
Additionally, this mapping provides a structured representation regarding the applicability of a defence mechanism for a particular attack based on any of the psychological principals. Since knowledge is a fundamental requirement which is exerted in the dimensions attitude and behaviour, it is relevant for all principles.

Dimension		Psychological Principle/ Defence Mechanism	Authority	Social Proof	Liking, Similarity, Deception	Commitment, Reciprocation, Consistency	Distraction
Knowledge	**Attitude**	Persuasion Knowledge					
		Forewarning					
		Attitude Bolstering	■				■
	Behaviour	Reality Check	■	■			■
		Inoculation	■	■			■
		Decision Making					

Table 4: Mapping of defence mechanisms against attacks based on psychological principals. Grey illustrates applicability of a defence mechanism, while black indicates non-applicability

As visualized above, there seem to exist two kinds of attacks based on the psychological principles. Firstly, attacks that are mainly defendable through the dimension of attitude, namely authority, social proof and distraction. Secondly, attacks that require a training of both dimensions, attitude and behaviour, in particular attacks based on liking, similarity, deception and commitment, reciprocation, consistency.

We first consider the dimension of attitude. Persuasion knowledge increases the likelihood of perceiving manipulation or deception attempts. Hence, it is relevant against all attack principles. In particular, since the main goal of social engineering attacks is to manipulate and influence the victims. In the same manner, forewarning is also relevant against all attacks based on the named psychological principles. Due to the fact that an attacker generally attempts to induce a pressure situation to his victim this mechanism generates awareness towards the malicious intentions. Especially, if forewarnings are combined with a precedent training of persuasion knowledge, the forewarning might trigger the recognition of attacks. Attitude bolstering is suitable for attacks based on Social Proof, Liking, Similarity, Deception, and Commitment, Reciprocation, Consistency if the security policies of the organisation are setup well. Mainly, because attacks based on this principles try to exploit a positive relationship built before the attack and/or try to assemble pressure due to societal obligations. However, the aim is to provoke the victim to practice a noncompliant security behaviour. By strengthening the attitude and improving awareness that not following the security policies can be harmful, attacks may be prevented. Given that the security policies are setup in a proper way. Attitude bolstering does not work well for attacks based on distraction and authority principles, since the main reason those attacks succeed is not that the victim is intentionally violating security policies. If those attacks

succeed, the user is either not aware that he is violating a policy or he is acting in good faith, obeying orders. The latter, is a fundamental principle of most organisational structures and therefore it would be risky to challenge this behaviour in a large scale. We discuss this idea in more detail at the end of this section.

Moving on to the dimension of behaviour, reality checks and inoculation are not applicable to attacks based on authority, because our societal system is based on and functions through authority. If a social proof is coherent, it just underlines how we function as a social being. And attacks based on distraction may not be countered, because limited attention is a characteristic that is not changeable. However, reality checks and inoculation are relevant to attacks based on liking, similarity, deception and commitment, reciprocation, consistency due to common unawareness that "naivety" in relationships and societal obligations is misused for this kind of attacks. Decision making is relevant to all kind of attacks since all attacks aim to influence the way the victim is making it's decisions, e.g. by not letting the victim think and getting him to rely on a heuristic. The defence mechanism helps in improving the victim's decision style or at a minimum evoke an awareness that decisions do not have to follow the first intuitive, impulsive reaction.

Another dimension we need to briefly mention is that the company should still preserve some kind of cooperative environment. Users should not overreact because they are afraid of being attacked. They still need to trust their colleagues to allow collaboration. Thus, another challenge to the user is to permanently do trade-offs between collaborating with his/her colleagues and avoiding/countering attacks. This often involves not following a policy for practical purposes, especially if they are contradictory and/or badly designed. As this kind of trade-off is very challenging to users and bears the risk that users are overburdened and simply give up in either of the two directions, as shown by Adams and Sasse (1999) in regards to user passwords.

8. A Gap Analysis of Missing Defence Mechanisms in IT Security against Social Engineering

As indicated above, the available defence mechanisms can be classified into the dimensions attitude and behaviour, which in turn exert knowledge. Table 5 presents a mapping of defence mechanisms comparing suggestions in IT security against techniques known in social psychology. When comparing the dimension attitude, the limited scope of IT security becomes evident. As established in Section 4, in the dimension attitude IT security considers establishment of policy and security awareness programs. The purpose of security awareness programs is twofold. Firstly, it is concerned with getting users to know and adhere to the established policy. Secondly, security awareness program's scope is usually limited to the provision of basic knowledge about social engineering. In comparison social psychology offers distinctively more. Although some approaches may be at least partly covered. Forewarning can be seen as included in the education of social engineering basics, as malicious intention of social engineers certainly belongs to basic knowledge about social engineering. But persuasion knowledge goes beyond social engineering basics as it includes knowledge about persuasion strategies as well as counter tactics to rely

on in any persuasive situation. For reliance on attitude bolstering good knowledge about security policy is necessary. Again IT security does the first step in user education, but fails in the second step, the enhancement of this knowledge. The use of attitude bolstering, implies not only the knowledge about policy but its implications and a thought process initiated by each user that strengthens his attitude to e.g. keep sensitive information private. The necessity to perform a reality check can directly be deduced from the concept of 'optimism bias', as illustrated in Section 3. It might partially be covered in security awareness programs. A reality check might be done for e.g. spam mails. But as this particular reality check has a technical background and people tend to dismiss their possible failure by it being a technical detail and in the same time greatly underestimating personal susceptibility, it is important to demonstrate to them their failure in a non-technical environment as well.

Dimension		**IT Defence Mechanisms**	**Psychological Defence Mechanisms**
Knowledge	**Attitude**	Policy Compliance	-
		Security Awareness Program	Forewarning
		-	Persuasion Knowledge
		-	Attitude Bolstering
		-	Reality Check
	Behaviour	Audit	-
		-	Inoculation
		-	Decision Making

Table 5: Comparison of defence mechanisms suggested in IT security and social psychology

Comparing mechanisms in Table 5 presents another crucial finding. The dimension behaviour is under-represented in IT security. The only suggestion made for this dimension is to verify correct behaviour via audits. But IT security fails to actually enhance secure behaviour. Training correct behaviour as part of security awareness programs is, as indicated in Section 4, recommended by only a few authors and is usually at most done for spam mails. Even though this is the application of inoculation, this is only one possible social engineering attack and a particularly technical one as well. Focus should again also be set on the persuasive nature of social engineering attacks. Hence trainings could for example include role plays. Additionally, it has been proven effective to alter the decision making process by conducting decision trainings where users make a "similar" decision in various appearances.

9. Discussion and Future Work

Previously, we have discussed (i) a mapping between defence mechanisms against attacks based on psychological principals and we identified (ii) gaps in IT security. Firstly, we want to elaborate on our findings regarding (i). While we provided a complete mapping, we are aware that it may be regarded as subjective. But as far as we are aware, this is the best structured comparison available. Furthermore, it is based on the results of our literature review and bears no experimental validation. To improve the mapping and furtherly validate it, we plan on conducting studies based on e.g. inoculation trainings to measure its influence regarding the psychological principles generally and particularly regarding the principles liking, similarity, deception and commitment, reciprocation, consistency. In a first step, we proposed a serious game (Beckers and Pape, 2016; Beckers et al., 2016) that helps players to understand how social engineering attacks work. The game can be played based on the real scenario in the company/department or based on a generic office scenario with personas that can be attacked. Our game trains people in realizing social engineering attacks in an entertaining way, which shall cause a lasting learning effect. In a next step we want to evaluate the collected data for further validation.

Secondly, we want to discuss the results regarding (ii). As indicated, both dimensions, attitude and behaviour, are represented inadequately in IT security when compared to recommendations from social psychology. To counter this gap, we envision two strategies for available security awareness programs (as shown in Table 5).
In a first strategy persuasion resistance trainings should be conducted.
They should include a broad approach to social engineering including psychological principles and their effects, possible counter strategies, the initiation of attitude bolstering. As optimism bias is a strong enabler of successful social engineering, it would be desirable to demonstrate users their susceptibility. This step is particularly promising, as it is feasible with little monetary effort. The second strategy is persuasive situation role plays. It is conceivable to include experiential exercises in this step as well as repeated decision trainings that force users to re-evaluate their knowledge and attitude by making a "similar" decision multiple times. This step is more effortful and it might suffice to only educate key personnel as it includes "live" training sessions guided by possibly costly trainers, actors or generally personnel capable of creating persuasive situations.

Dimensions	Future defence strategies
Attitude	Persuasion resistance training
Behaviour	Persuasive situation role plays

Table 6: Envisioned training strategies as part of security awareness

Additionally, it is worth to bear in mind, that although it is desirable to educate staff, there possibly exists a fine line to not overwhelm users with rules and knowledge.

10. References

Adams, A. and Sasse, M.A., 1999. Users Are Not the Enemy. *Commun. ACM*, 42(12), pp.40–46. Available at: http://doi.acm.org/10.1145/322796.322806.

Anon, 2011. Dimensional Research Study about Social Engineering. In *Analysis of Social Engineering Threats with Attack Graphs*. Beckers, Kristian Krautsevich, Leanid Yautsiukhin, Artsiom.

Bakhshi, T., Papadaki, M. and Furnell, S., 2008. A Practical Assessment of Social Engineering Vulnerabilities. In N. L. Clarke & S. Furnell, eds. *2nd International Conference on Human Aspects of Information Security and Assurance, {HAISA} 2008, Plymouth, UK, July 8-9, 2008. Proceedings*. University of Plymouth, pp. 12–23. Available at: http://www.cscan.org/openaccess/?paperid=53.

Barrett, N., 2003. Penetration testing and social engineering: hacking the weakest link. *Information Security Technical Report*, 8(4), pp.56–64.

Beckers, K. and Pape, S., 2016. A Serious Game for Eliciting Social Engineering Security Requirements. In *RE*.

Beckers, K., Pape, S. and Fries, V., 2016. HATCH: Hack and Trick Capricious Humans - A Serious Game on Social Engineering. In *BCS HCI*.

Brehm, J.W., 1966. *A theory of psychological reactance*, New York: Academic Press.

Briñol, P., Rucker, D.D. and Petty, R.E., 2015. Naïve theories about persuasion: Implications for information processing and consumer attitude change. *International Journal of Advertising*, 34(1), pp.85–106.

Butavicius, M. et al., 2015. Breaching the Human Firewall : Social engineering in Phishing and Spear-Phishing Emails. *Australasian Conference on Information Systems*, pp.1–11.

Cacioppo, J.T. et al., 1986. Central and peripheral routes to persuasion: An individual difference perspective. *Journal of Personality and Social Psychology*, 51(5), pp.1032–1043.

Cialdini, R.B., 2009. *Influence: the psychology of persuasion* EPub editi., New York: Collins.

Ferreira, A., Coventry, L. and Lenzini, G., 2015. Principles of Persuasion in Social Engineering and Their Use in Phishing. In T. Tryfonas & I. Askoxylakis, eds. *Human Aspects of Information Security, Privacy, and Trust SE - 4*. Lecture Notes in Computer Science. Springer International Publishing, pp. 36–47. Available at: http://dx.doi.org/10.1007/978-3-319-20376-8_4.

Frangopoulos, E.D., Eloff, M.M. and Venter, L.M., 2012. Psychosocial Risks: can their effects

on the Security of Information Systems really be ignored? In N. L. Clarke & S. Furnell, eds. *6th International Symposium on Human Aspects of Information Security and Assurance, {HAISA} 2012, Crete, Greece, June 6-8, 2012. Proceedings*. University of Plymouth, pp. 52–63. Available at: http://www.cscan.org/openaccess/?paperid=35.

Fransen, M.L. et al., 2015. Strategies and motives for resistance to persuasion : an integrative framework. *Frontiers in psychology*, 6(August), pp.1–12.

Friestad, M. and Wright, P., 1994. The Persuasion Knowledge Model: How People Cope with Persuasion Attempts. *Journal of Consumer Research*, 21(1), pp.1–31. Available at: http://www.jstor.org/stable/2489738.

Gigerenzer, G. and Todd, P.M., 1999. *Simple Heuristics that Make us Smart*, Oxford: Oxford University Press.

Gilbert, D.T., 2002. Inferential correction. In T. Gilovich, D. Griffin, & D. Kahneman, eds. *Heuristics and biases*. New York: Cambridge University Press.

Gouldner, A.W., 1960. The Norm of Reciprocity: A Preliminary Statement. *American Sociological Review*, 25(2), pp.161–178. Available at: http://www.jstor.org/stable/2092623.

Gragg, D., 2003. A multi-level defense against social engineering. *SANS Reading Room, March*, 13.

Gulati, R., 2003. The Threat of Social Engineering and your defense against it. *SANS Reading Room*.

Ivaturi, K. and Janczewski, L., 2011. A Taxonomy for Social Engineering attacks. *Proceedings of CONF-IRM*.

Jones, C., 2004. Understanding and Auditing. *SANS Institute Infosec Reading room*.

Kahneman, D., 2003. A perspective on judgment and choice: mapping bounded rationality. *The American psychologist*, 58(9), pp.697–720.

Landman, J. and Petty, R., 2000. "It Could Have Been You": How States Exploit Counterfactual Thought to Market Lotteries. *Psychology & Marketing*, 17(4), pp.299–321.

Lott, A.J. and Lott, B.E., 1965. Group Cohesiveness as Interpersonal Attraction: A Review of Relationships with Antecedent and Consequent Variables. *Psychological Bulletin*, 64(4), pp.259–309.

Lydon, J., Zanna, M.P. and Ross, M., 1988. Bolstering Attitudes by Autobiographical Recall: Attitude Persistence and Selective Memory. *Personality and Social Psychology Bulletin*, 14(1), pp.78–86. Available at: http://psp.sagepub.com/content/14/1/78.abstract.

Lynn, M., 1989. Scarcity effects on desirability: Mediated by assumed expensiveness? *Journal of Economic Psychology*, 10(2), pp.257–274.

Manske, K., 2009. An Introduction to Social Engineering. *Information Security Journal: A Global Perspective*, 9(5), pp.1–7.

Milgram, S., 1963. Behavioral study of obedience. *The Journal of Abnormal and Social Psychology*, 67(4), p.371.

Milgram, S., 1974. *Obedience to authority*, London: Tavistock.

Miller, C.H. et al., 2013. Boosting the Potency of Resistance: Combining the Motivational Forces of Inoculation and Psychological Reactance. *Human Communication Research*, 39(1), pp.127–155.

Mitnick, K.D. and Simon, W.L., 2011. *The art of deception: Controlling the human element of security*, John Wiley & Sons.

Petty, R.E. and Cacioppo, J.T., 1996. *Attitudes and persuasion: Classic and contemporary approaches*, Boulder, CO, US: Westview Press.

Pfleeger, S.L., Sasse, M.A. and Furnham, A., 2014. From Weakest Link to Security Hero: Transforming Staff Security Behavior. *Journal of Homeland Security and Emergency Management*, 11(4), pp.489–510.

Rusch, J.J.J., 1999. The " Social Engineering " of Internet Fraud. *nternet Society's INET'99 conference*, pp.1–12. Available at: http://www.isoc.org/isoc/conferences/inet/99/proceedings/3g/3g_2.htm.

Sagarin, B.J. et al., 2002. Dispelling the illusion of invulnerability: The motivations and mechanisms of resistance to persuasion. *Journal of Personality and Social Psychology*, 83(3), pp.526–541. Available at: http://doi.apa.org/getdoi.cfm?doi=10.1037/0022-3514.83.3.526.

Scheeres, J.W., 2008. *Establishing the human firewall: reducing an individual's vulnerability to social engineering attacks*,

Schneier, B., 2008. The Psychology of Security. In S. Vaudenay, ed. *Progress in Cryptology – AFRICACRYPT 2008 SE - 5*. Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 50–79. Available at: http://dx.doi.org/10.1007/978-3-540-68164-9_5.

Stajano, F. and Wilson, P., 2011. Understanding Scam Victims: Seven Principles for Systems Security. *Commun. ACM*, 54(3), pp.70–75. Available at: http://doi.acm.org/10.1145/1897852.1897872.

Stanovich, K.E. and West, R.F., 2002. Individual differences in reasoning: Implications for the rationality debate. In T. Gilovich, D. Griffin, & D. Kahneman, eds. *Heuristics and biases*. New York: Cambridge University Press.

Thornburgh, T., 2004. Social Engineering: The "Dark Art." In *Proceedings of the 1st Annual Conference on Information Security Curriculum Development*. InfoSecCD '04. New York, NY, USA: ACM, pp. 133–135. Available at: http://doi.acm.org/10.1145/1059524.1059554.

Verizon, 2012. Data Breach Investigations Report. Available at: http://www.verizonenterprise.com/resources/reports/rp_data-breach-investigations-report-2012-ebk_en_xg.pdf [Accessed January 13, 2016].

Verizon, 2013. Data Breach Investigations Report. Available at: http://www.verizonenterprise.com/resources/reports/rp_data-breach-investigations-report-2013_en_xg.pdf [Accessed January 13, 2016].

Veseli, I., 2011. *Measuring the Effectiveness of Information Security Awareness Program*. Gjøvik University College.

Warkentin, M. and Willison, R., 2009. Behavioral and policy issues in information systems security: the insider threat. *European Journal of Information Systems*, 18(2), pp.101–105.

Weinstein, N.D., 1980. Unrealistic Optimism About Future Life events. *Journal of Personality and Social Psychology*, 39(5), pp.806–820.

Winkler, I.S. and Dealy, B., 1995. Information Security Technology?...Don't Rely on It A Case Study in Social Engineering. In *Fifth Usenix Security Symposium*. pp. 1–6.

Xu, A.J. and Wyer, R.S.J., 2012. The Role of Bolstering and Counterarguing Mind-Sets in Persuasion. *Journal of Consumer Research*, 38(5), pp.920–932. Available at: http://www.jstor.org/stable/10.1086/661112.

A.5 A Structured Comparison of Social Engineering Intelligence Gathering Tools

Kristian Beckers, Daniel Schosser, Sebastian Pape, and Peter Schaab. A structured comparison of social engineering intelligence gathering tools. In *Trust, Privacy and Security in Digital Business - 14th International Conference, TrustBus 2017, Lyon, France, August 30-31, 2017, Proceedings*, pages 232–246, 2017. doi: 10.1007/978-3-319-64483-7_15. URL https://doi.org/10.1007/978-3-319-64483-7_15

A Structured Comparison of Social Engineering Intelligence Gathering Tools

Kristian Beckers[1(✉)], Daniel Schosser[1], Sebastian Pape[2], and Peter Schaab[1]

[1] Institute of Informatics, Technische Universität München (TUM), Boltzmannstr. 3, 85748 Garching, Germany
beckersk@in.tum.de

[2] Faculty of Economics and Business Administration, Goethe University Frankfurt, Theodor-W.-Adorno-Platz 4, 60323 Frankfurt, Germany

Abstract. Social engineering is the clever manipulation of the human tendency to trust to acquire information assets. While technical security of most critical systems is high, the systems remain vulnerable to attacks from social engineers. Traditional penetration testing approaches often focus on vulnerabilities in network or software systems. Few approaches even consider the exploitation of humans via social engineering. While the amount of social engineering attacks and the damage they cause rise every year, the defences against social engineering do not evolve accordingly. However, tools exist for social engineering intelligence gathering, which means the gathering of information about possible victims that can be used in an attack. We survey these tools and present an overview of their capabilities. We concluded that attackers have a wide range of intelligence gathering tools at their disposal, which increases the likelihood of future attacks and allows even non-technical skilled users to apply these tools.

Keywords: Social engineering · Threat analysis · Security awareness · Security tools

1 Introduction

"The biggest threat to security of a company is not a computer virus, an unpatched hole in a key program or a badly installed firewall. In fact, the biggest threat could be you [...] What I found personally to be true was that it's easier to manipulate people rather than technology [...] Most of the time organizations overlook that human element". These words from Kevin Mitnick [7] were made over a decade ago and are still of utmost importance today.

As security technology improves the human user remains the weakest link in system security. It is widely accepted that the people of an organization are therefore both the main vulnerability of any organization's security as well as the most challenging aspect of system security [6,27]. Hadnagy [17] defines social engineering as "Any act that influences a person to take an action that may or

J. Lopez et al. (Eds.): TrustBus 2017, LNCS 10442, pp. 232–246, 2017.
DOI: 10.1007/978-3-319-64483-7_15

may not be in their best interest". Numerous security consultants consider it a given for themselves as well as for genuine attackers to access critical information via social engineering [14,43].

The harm of social engineering attacks has been discussed in various reports. In 2003 Gulati [15] reported that cyber attacks cost U.S. companies $266 million every year and that 80% of the attacks are a form of social engineering. Although not being very recent assessments of the situation, it seems that little has changed until today. A study of 2011 from Dimensional Research [9] shows that nearly half of the considered large companies and a third of small companies fell victim of 25 or more social engineering attacks in the two years before. The study further shows that costs per incident usually vary between $25 000 and over $100 000. Furthermore, surveys, like Verizon's *Data Breach Investigation Report* [41,42] show the impact of social engineering. According to these studies the impact has grown from 7% of breaches in 2012 to 29% of breaches in 2013. These numbers should not be ignored and active support for mitigating these threats is needed.

Even though companies are aware of the social engineering problem, they have little tools available to even assess the threat for themselves. Hiring penetration testing companies that *attack* their clients and show weaknesses in their defences is one available option. However, these tests have a number of inherent problems. Particularly, to address legal issues high effort has to be invested upfront [44]. In addition, the test outcome is closely related to the limited scope of the test. A tester may find that some employees are violating security policies. While this is an important finding that lets a company improve the education of their employees, the completeness of these kind of tests is an issue. Only few employees can be tested on only few occasions. Moreover, experiments indicate that this approach is difficult, due to humans' demotivation when confronted with these testing results [10].

A number of tools are available that enable intelligence gathering. On one side using these tools a social engineer can gather information that help him attack persons or organizations. On the other side, these tools provide an organization with an excellent alternative to pen testing or awareness trainings, as they allow to analyse possible vulnerabilities. However, a structured survey on the tools' capabilities is missing so far.

We believe to improve the current situation by conducting a structured survey of social engineering intelligence gathering tools and contribute the following:

- A classification of existing tools regarding categories such as proposed purpose, price, perceived usability, visualization of results etc.
- A survey of information types retrieved by the tools regarding information about company employees and their communication channels, as well as related information e.g. company policies;
- A discussion of how even simple attacker types can use these tools for sophisticated social engineering attacks.

The remainder of our paper is organised as follows. Section 2 outlines the criteria for comparison, and Sect. 3 presents the results of our comparison. Section 4 concludes and provides directions for future research.

2 Social Engineering Basics and Tool Criteria

We acquire a basic understanding of social engineering and the general process attackers follow in Sect. 2.1. During the process various information is gathered about people, whom social engineers attack. Section 2.2 details our categorization of this *social engineering information* based on related work. Furthermore, we classify the tools on their *potential of applicability*, which describes the barriers that may or may not prevent an attacker from using them. For example, a tool that has a high price and poor usability will have little potential to be used by any attacker.

2.1 The Social Engineering Process

Various works report an underlying process to social engineering [17,21,27], which have recently been unified by Milosevic [26]. A social engineering attack consists of multiple phases as summarized in Table 1. In phase one the attacker conducts surveillance to identify a person within the inner circle of the targeted company. This person shall have access to the information the attacker desires. The next phase focuses on finding out as much about this person as possible. Every bit of information can help the attacker to manipulate the victim and her trust. During the pretexting phase the attacker starts building a relationship to the victim. Afterwards the attacker exploits the built up trust in the relationship and evaluates the gathered information in the post-exploitation phase.

Table 1. Overview of social engineering phases by Milosevic [26]

Phase	Description
Pre-engagement interactions	Find targets with sufficient access to information/knowledge to perform an attack
Intelligence gathering	Gather information on each of the valid targets. Choose the ones to attack
Pretexting	Use gathered information to build a relationship to the target. Gain victims' trust to access additional information
Exploitation	Use the built up trust to get the desired information
Post-exploitation	Analyze the attack and the retrieved information. If necessary return to a previous phase to continue the chain of attack until the final information has been retrieved

2.2 Social Engineering Information

This section focuses on types of information that can be gathered by a tool, in the following referred to as criteria. All criteria cover one or more essential information for social engineering attackers. The more criteria a tool covers, the more interesting it is for a social engineer during information gathering.

Communication Channels. Communication channels are one of the most relevant information for a social engineer. This category will list which channels can be found by a certain tool. Possible channels are "Telephone Numbers", "Social Media Accounts", "E-Mails", "Instant Messengers", "Friends", "Personal Information" and "Private Locations" [23,27].

User Credentials. Some tools have access to databases which contain leaked user credentials. If a social engineer gets access to login information of a certain employee, it simplifies the conduction of an attack. Firstly, he can directly access a victim's accounts. Secondly, the attacker could pose as someone else, e.g. an administrator from the IT department, and by having access to the target's data convince his victim to act in a certain way [18,27].

Locations. Some tools are especially designed to gather location data, while others provide them as a byproduct. Both, work addresses as well as an employee's private addresses can be useful for multiple purposes. Location data can be gathered from social media as it is embedded in photos and videos taken by cellphones. Also posts on social media can be tagged with a location. Other tools can convert IP addresses into physical locations and therefore find the physical locations of technical equipment [18,35].

Job Positions. By retrieving the job position of an employee the social engineer can figure out what kind of information someone has access to. Based on job title, the attacker can draw conclusions about whether an employee is new to a company, what the hierarchy within the company looks like and much more. Based on the organization's structure, it is possible to use techniques such as name-dropping, using the name of someone higher in the company's hierarchy, to pressure the target into revealing information [18,27].

Company Lingo. One of the easiest ways to convince someone of being authorized to access some information is by knowing the correct lingo [27]. Lingo means the words and abbreviations employees use within a company. Although this information is of great importance, it is very challenging to get access to. Knowledge about the lingo can be obtained by getting access to company manuals, internal reports or talking to employees.

Personal Information. The more personal information an attacker has on his target, the easier it is to find the correct angle and pressure points. One example would be well-defined spear-phishing e-mails using a person's interests. In case the e-mail contains enough private information to make it believable, the target is far more likely to open an attachment [19,35].

2.3 Potential for Applicability

This section presents the evaluation criteria to generally classify the software.

Proposed Purpose. Some of the tools are actually designed to gather information on a person or company in the context of social engineering. However, a user can also use tools for attacks which were designed for something completely different than social engineering.

Price. While some tools are free, others can be quite expensive and therefore might not be applicable for a quick self assessment. In some cases the tool itself is free, but for some features the user needs to have an API key that can be costly. This criteria focuses on the prices of each tool and its limitations coming with different price tiers.

Usability. Based on the user interface and the amount of documentation provided, this category assesses the ease of usage. The underlying question is if the usability of a tool allows a company to perform its own threat assessment.

Input Parameters. Some tools have a broad range of possible search arguments, but most tools need specific information to initiate a search. Depending on which specific piece of information is required by the tool, this might limit the social engineer in the decision what tools to use.

Visualize Output. Some tools print all information into tables while others have better ways of visualizing gathered information. For example location data can be illustrated by marking the positions on a map, instead of only providing GPS coordinates.

Ranking of Results. As the amount of gathered information grows, the more valuable an adequate selection and sorting becomes. Therefore, filtering irrelevant information is helpful in focusing on more promising targets/information. We did not find significant support for filtering in the analysed tools and therefore do not list this criteria in the following.

Suggesting Counter-Measures. Most of the tools are only designed to gather information and do not inform how to protect this information. While this is not relevant for social engineers, it is highly relevant for those who want to protect themselves against attackers and against information gathering in general. Note that none of the tools suggest countermeasures, therefore we did not list the category in the following.

3 Comparing Social Engineering Tools and Webpages

In the following section, we introduce and analyze relevant tools and webpages. In a second step we provide an overview over the types of information that can be gathered by them.

3.1 Social Engineering Tools and Webpages

We compiled the following list of social engineering tools by using the following words "social engineering and tool or application or script or webpage" in a Google[1] search and the list published by Hadnagy [17]. Three security researchers analysed the results independently and we included all tools and webpages that

[1] https://www.google.de.

they agreed on having the potential to help a social engineering attacker conduct the process outlined in Sect. 2.1. We identified the following tools and webpages that met our criteria.

Maltego (Kali Linux Edition, Version 3.6.1). Maltego [32] is an intelligence and forensics application. Before starting a search, the user can choose between different machines. Every machine has its own purpose and is designed for a particular attack vector. Maltego offers 12 default machines within the software such as: *Company Stalker* This machine tries to get all e-mail addresses at a domain to resolve them on social networks. It also gets documents and extracts meta data. As an input, it needs a company's domain. *Find Wikipedia Edits* This machine takes a domain and looks for possible Wikipedia edits. *Footprint L1* This module performs a level 1 (fast, basic) footprint of a domain. *Person - E-Mail Address* This machine tries to obtain someone's e-mail address and checks where it's used on the internet.

Maltego combines multiple modules to gather information from various sources and represents them in an easy to understand way in form of a bubble diagram. The user can start of with a domain name, a username, an IP address or the name of a person depending on which module he wishes to use. The gained information can be used for further research e.g. as input for other modules.

Recon-ng (Version 4.8.0). Recon-ng [40] is a full-featured Web Reconnaissance framework. It is based on a large list of modules which can be used to gather information about a specific target. The modules range from host information to social media. The user is free to chain these modules after each other and by starting with a single domain name, the database can be filled with employee names, their e-mail addresses, usernames, passwords and geolocations of all involved servers. The final reports can be exported in json, csv, xml, html or as a pdf. Similar to the Social Engineering Toolkit and Metasploit its user interface is console based.

Cree.py (Version 1.4). Cree.py [20] is a geolocation Open Source Intelligence (OSINT) tool. It is designed to gather geolocation related information from online sources like social networks. This information can be filtered by location or date and is presented on a map. Therefore, Cree.py is useful to follow the trace of where a person has been over the time of using certain social media platforms. Examples would be Instagram, Twitter or Tumblr which gather location data on where photos or posts have been created. These information can be displayed on a map and recreate a trace of places where a person has been.

Spokeo. Spokeo [38] is a search engine for people in the United States of America. There exist equivalent versions for other countries e.g. Pipl.com and PeekYou.com index people from all over the world. By entering the name, e-mail address, phone number, address or username of a person all related people matching the provided criteria are reported back. Depending on the wanted detail of the provided report, the price varies.

Social Engineering Toolkit (SET). SET [16] does not focus on finding information about a person. SET rather uses information on persons to e.g. send them phishing e-mails or gather information about company networks. The SET allows integration with other tools such as Metasploit that contain various scripts for vulnerability testing.

The Wayback Machine. The Wayback Machine [39] is an archive of the internet. The vendor claims to provide the history of more then 427 billion web pages (as of July 2015). The platform creates snapshots of websites and allows a user to go back to older versions of a website that have been replaced by newer ones.

theHarvester (Version 2.7). The Harvester [12] is designed to gather e-mail addresses, subdomains, hosts, and open ports from public sources. These sources contain search engines, PGP key servers and the SHODAN [36] computer database for internet-connected devices.

Whitepages. The *Whitepages* [5] website supports persons in finding people, their addresses and telephone numbers, private and from work. The service focuses on the U.S. and also provides reverse phone searches and similar means to identify a person based on technical information such as a phone number.

Background Checks. The *freebackgroundcheck.com* [1] website provides information about people that has been collected by background checks on them for e.g. a telecommunication provider. The intention is that people can get informed what information is available about them and most likely checked in situations such as job interviews. The website *Instant Checkmate* [2] on the other hand focuses on providing information to the public about people's arrest records and criminal behaviour.

Tax Records. Especially in the United States it is very easy to gain access to government information, as most data is publicly available [30]. Every person interested in the data can get access to arrest records, tax records and more for a small monetary fee per request. In addition, Ratsit in Sweden [34], Veroposi in Finland [4], Skatterlister in Norway [3] and recently the Federal Board of Revenue in Pakistan [31] also publish tax records online.

Company Related Information. As social engineers thrive to know as much about the social surroundings of a target as possible, there are a lot of tools, that help gathering social related information about a target. Websites like *KnowEm* [22] and *Namechk* [29] allow to search on more then 600 social media networks, if a username is already allocated or still available. While this is not the primary purpose of the website, an attacker can use this to track down social media networks, which a target is using. *SocialMention* [37] is a platform, that searches for user-generated content like posts, blogs, videos, etc. from a specific user. By gathering this kind of information the attacker learns a lot about the target and his behavior.

In most cases a social engineer is not after private information about a target, but work related information. This is due to an attacker generally trying to get access to work related sensitive information. Websites such as *Monster* [28],

LinkedIn [24] and *Xing* [45] are good sources for collecting CVs and current job positions of people related to the target. In addition platforms like *career-builder* [8] and *glassdoor* [13] provide information about open job offers and expected earnings. *Hoovers* [11], *MarketVisual* [25] and *LittleSis* [33] are useful to gain knowledge about the social networks of employees. Especially for larger companies, these websites offer information about who is connected to whom.

3.2 Analyzing the Social Engineering Attack Potential

After having established each tool's characteristics, it is important to know, what tool is able to retrieve which kind of information. Some tools are able to collect more information than others and some information can only be found with a specific tool. Table 2 provides an overview of the tools survey. Furthermore, Table 3 provides a refinement of the previous table considering the potential for applicability categories introduced in Sect. 2.3 for selected tools and webpages. For space reasons we do not show the information for all tools and websites.

Our goal is to show the utility of these tools for attackers. Therefore, we selected three attack types mentioned repeatedly [17,23,27]: *Phishing*, *Baiting*, and *Impersonation*. We describe these below including their needs of two essential information categories: *communication channels* and *company knowledge*. An attacker requires communication channels since the attacker has to communicate with a victim to exploit her trust. In addition, an attacker requires knowledge about the company to know whom to attack and how to get the companies employees' trust. The more details an attacker knows, the more likely people

Table 2. Social engineering tool comparison

	SET	Maltego	Recon-ng	Cree.py	Spokeo	Wayback Machine	theHarvester	knowem.com	Whitepages	Instant Checkmate	freebackgroundcheck.org
Search by Person/ Company	o	++	++	++	++	++	++	+	++	++	++
Retrieve E-Mail Address	o	++	++	o	o	o	++	o	o	o	o
Retrieve Username/ Password	o	o	++	o	o	o	o	o	o	o	o
Retrieve Job-Title	o	o	++	o	o	o	o	o	o	++	++
Retrieve Locations	o	+	+	++	+	o	o	o	++	++	++
Retrieve Personal Data	o	o	o	o	++	o	o	+	+	++	++
Usability	+	+	+	++	++	++	+	++	++	++	++
Visualize Output	+	++	+	++	++	++	+	++	++	++	++
Retrieve Company Lingo	o	o	o	o	o	o	o	o	o	o	o
Free to use	++	++	++	++	o	++	++	++	++	o	o

o Does not apply or cannot be used in this case
\+ Does apply in some cases, does collect limited information
++ Does fully apply, does gather the amount/quality of information needed

Table 3. Potential for applicability

Category	Maltego	Recon-ng	Cree.py	Spokeo	The wayback machine	The Harvester
Proposed purpose	Delivery of a threat picture of an organization's environment	Enables conduction of web-based reconnaissance	Provision of geolocation related information from social media	Provision of personal information	Archive for webpages and other media	Gather e-mails, subdomains, hosts and open ports from different public sources
Price	Free community edition, Full license $760first year, $320additional year	Free, API Keys up to $60,000	Free	Free basic information, $4.95month for detailed reports, $9.95 for court records	Free	Free
Usability	Easy to understand UI. Basic knowledge about structure and connection of information and available machines required	Terminal based tool. Basic knowledge about structure and connection of information and available modules required	Easy to use due to UI and step by step guidance	Easy to use due to step by step guidance	Easy to use due to centralization in single search field	Terminal based tool. Simple execution
Input parameters	Depending on the machine name, web domain, username, company name	Depending on the module domain name, URL, name	Username	Name, e-mail, phone, username, address	Web domain	Company name, web domain
Visualize parameters	Bubble diagram. Color coded data categories. Bubble sizes according to data amount	Local database exportable to various formats	Data listed, pins on map	Pins on map	Calendar based data entries. Available snapshots highlighted	Data tables
Relevant phases	Phase 1 - Pre-Engagement Interactions, Phase 2 - Intelligence Gathering	Phase 2 - Intelligence Gathering	Phase 2 - Intelligence Gathering	Phase 1 - Pre-Engagement Interactions, Phase 2 - Intelligence Gathering	Phase 1 - Pre-Engagement Interactions, Phase 2 - Intelligence Gathering	Phase 2 - Intelligence Gathering

believe he has a relation to the company. We detail these information needs for the attack types below and refine them in Table 4.

Table 4. Mapping of social engineering characteristics to attack types

	Attack type		
	Phishing	Baiting	Impersonation
Communication			
Telephone number	x		
Friends	x		x
Personal information	x		x
Private locations	x		x
EMail	x		
Instant messenger	x		
Co-workers: communication			x
Company knowledge			
Co-workers: new employee			x
Co-workers: hierarchies			x
Lingo	x		x
Facilities: security-measures		x	x
Facilities: company location		x	x
Websites	x		
Policies: software		x	
Policies: network		x	
Policies: organization		x	

Phishing refers to masquerading as a trustworthy entity and using this trust to acquire information or manipulating somebody to perform an action. This often appears in an unguided way via email to thousands of possible victims. Recently, spear-phishing attacks happen, which aim for a specific target instead of the broader mass. The social engineer gathers as much intelligence about the target as he can or needs and then prepares a tailored message for the victim.

Information Needs: Phishing attacks are mainly based on communicating with the victim, therefore the amount of information on communication channels is critical. The more channels an attacker has, the easier it is, to find one that can help bridge the gap between the engineer and the victim. In addition, the more company knowledge exist, the more targeted the attack can be.

Baiting is to leave a storage medium (e.g., a USB stick) inside a company location that contains malicious software (e.g., a key logger). The malicious software is executed automatically when the stick is inserted in a computer.

Table 5. Tool coverage for communication channels

	Cree. py	Gitrob	KnowEm	LinkedIn	Maltego	Namechk	Recon-ng	Spokeo	theHarvester	Wayback machine	Wire-shark	Xing
Telephone number							x					x
E-mail				x	x		x		x			x
Instant messenger			x		x	x		x				x
Friends			x	x	x	x						x
Personal information	x		x	x		x		x				x
Private locations	x							x				x

Table 6. Tool coverage for company knowledge

	Cree. py	Gitrob	KnowEm	LinkedIn	Maltego	Namechk	Recon-ng	Spokeo	theHarvester	Wayback Machine	Wire-shark	Xing
Company locations	x			x			x	x	x			x
Company lingo												
Special knowledge				x	x		x					x
New employees				x	x							x
Hierarchies				x	x							x
Websites					x		x		x	x		
Facility security measures		x									x	
Security policies		x							x		x	
Software policies		x					x				x	

Information Needs: Baiting is a passive attack vector, which does not need direct interaction with the victim. Therefore, the focus lies on gathering company knowledge. In particular, locations and walking routes of employees for placing the storage medium are essential.

Impersonation is to play the role of someone a victim is likely to trust or obey, e.g. an authority figure. The attacker fools the victim into allowing him access to the desired location or information. Usually, attackers prepare well for an impersonation and leverage vast amount of information.

Information Needs: For a successful impersonation attack company knowledge is a priority. The social engineer needs knowledge of numerous areas of the company. The more information he has on the persona he is playing, the more convincing he can be. Communication channels are of less importance, since the victim is approached in person.

We illustrate the degree to which the information needs of a social engineer can be covered for the discussed attack types. Tables 5 and 6 match tools with communication channels and company knowledge. Table 6 reveals that numerous tools cover information gathering for locations, websites, new employees etc. of

Table 7. Tools vs. AttackType knowledge with P for phishing, I for impersonation, and B for baiting

	Cree.py	Gitrob	KnowEm	LinkedIn	Maltego	Namechk	Recon-ng	Spokeo	theHarvester	Wayback machine	Wire-shark	Xing
Telephone number							P					P
Friends			P,I	P,I	P,I	P,I						P,I
Personal information	P,I		P,I	P,I		P,I		P,I				P,I
Private locations	P,I							P,I				P,I
E-mail				P	P		P		P			P
Instant messenger			P		P	P		P				P
Co-workers: communication												
Co-workers: newEmployee				I			I					I
Co-workers: hierarchies				I			I					I
Lingo	P,I	P,I	P,I	P,I	P,I	P,I	P,I	P,I	P,I	P,I	P,I	P,I
Facilities: security-measures		B,I									B,I	
Facilities: company location	B,I			B,I			B,I	B,I	B,I			B,I
Websites					P		P		P	P		

companies. However, the *Company Lingo* is not covered at all. Company lingo contains all abbreviations and specific terms used in a company and has been used by social engineers to bypass authentication mechanisms, e.g. personnel often thinks everyone knowing the company lingo belongs to the company [27].

For "Facility Security Measures", "Security Policies" and "Software Policies" there is a similar result. Besides *theHarvester* and *Recon-ng*, which can both only gather information concerning web-security like open ports or SSL-Encryption, all other tools are not directly suitable for social engineers. *Wireshark* needs physical access, which is not exactly what a social engineer prefers and *Gitrob* is one of the tools, with very slim chances of success. If the company has any security policies or hosts their sourcecode within the company, then *Gitrob* will most likely not be able to access it and therefore not gain any information.

To sum up, modern social engineers have a variety of tools at their disposal for information gathering, which they can use in numerous attacks. We provide an exemplary overview for phishing, baiting, and impersonation attacks and summarize in Table 7. The empty fields mean that three security researchers could not identify a use for that tool for any of the attacks above. Note that there are still some types of information that are difficult to gather for an attacker such as company lingo, but we have little doubt that in the future further tools and social media offers will fill this gap. Furthermore, our comparison showed that all tools have a good or great usability and provide easy to understand output. This means intelligence gathering can be used by an attacker with little technical

knowledge such as script kiddies. Therefore, we have to take the threats arising from increased and easily available knowledge for social engineering seriously.

4 Conclusions

We conducted a structured survey of social engineering tools, which ease the attacker's effort of finding information about victims. We mapped the information to their usefulness for phishing, impersonation or baiting attacks. Our analysis revealed that the social engineering threat is more dangerous than ever before, due to the number of tools at an attacker's disposal and the significant amount of detail they provide. We propose the following.

Implications for Possible Victims. People in general, not only employees in companies, can fall victim to social engineering. Therefore, people should find out what is available about them in the web using the tools or websites listed here. Ideally, stories of new contacts and unusual requests to secret information should be checked and verified more carefully than in the past. Means of protection can include false information released such a bogus address or non-existing hobbies. Any requests using this information identify possible social engineers.

Implications for Security Practitioners. Chief information officers and consultants should integrate a demonstration of the tools in this publication to raise awareness of the social engineering threat in companies. Just when people see the ease of collecting information with the tools and websites and how these are used e.g. in phishing, they can understand the need for strict security policies with regard to the release of data in the web.

Suggestions for Law Enforcement has to operate under the assumption that criminals will get all information about their victims without ever leaving their home or having mature computer skills. Everyone can be a social engineer and is a possible perpetrator. Countermeasures have to include network traffic analysis of how an attacker gathered the information for his attacks.

Limitations of the Tools. The only information type that social engineering tools do not provide today is the so-called *company lingo*, the abbreviations and specific words used in a company or domain. However, we are certain that in the future, tools combining machine learning and big data analysis will fill this gap.

Limitations of our Study. We conducted the study using a previous survey of tools and a web search engine. These sources can be extended in particular to including sites that are not indexed by web search engines e.g. in the dark web. This work will require a collaboration with a law enforcement agency.

Acknowledgements. This research has been partially supported by the Federal Ministry of Education and Research Germany (BMBF) with project grant number 16KIS0240.

References

1. Freebackgroundcheck. https://mybackgroundcheck.preemploy.com
2. Instant checkmate. https://www.instantcheckmate.com
3. Norwegian register. http://skattelister.no/
4. Tax information. http://www.veroporssi.com/
5. Whitepages. http://www.whitepages.com
6. Barrett, N.: Penetration testing and social engineering: hacking the weakest link. Inf. Secur. Tech. Rep. **8**(4), 56–64 (2003)
7. BBC News. How to hack people, October 2002. http://news.bbc.co.uk/2/hi/technology/2320121.stm
8. CareerBuilder. Job search engine. http://careerbuilder.com/
9. Dimensional Research. The risk of social engineering on information security, September 2011. http://docplayer.net/11092603-The-risk-of-social-engineering-on-information-security.html
10. Dimkov, T., van Cleeff, A., Pieters, W., Hartel, P.: Two methodologies for physical penetration testing using social engineering. In: Proceedings of ACSAC, ACSAC 2010, pp. 399–408. ACM (2010)
11. Dun & Bradstreet. Sales acceleration platform. http://www.hoovers.com/
12. Edge-Security. theHarvester. http://www.edge-security.com/theharvester.php
13. Glassdoor. Recruiting website. https://www.glassdoor.de/
14. Gragg, D.: A multi-level defense against social engineering. SANS Reading Room, 13 March 2003
15. Gulati, R.: The threat of social engineering and your defense against it. SANS Reading Room (2003)
16. Hadnagy. Social engineering toolkit (set). http://www.social-engineer.org/framework/se-tools/computer-based/social-engineer-toolkit-set/
17. Hadnagy, C.: Social Engineering: The Art of Human Hacking. Wiley, Indianapolis (2010)
18. Hadnagy, C.: The Official Social Engineering Portal (2015)
19. Internetsafety 101. Social Media Statistics (2013). http://www.internetsafety101.org/Socialmediastats.htm
20. Kakavas. Geolocation OSINT Tool. http://www.geocreepy.com/
21. Kee, J.: Social Engineering: Manipulating the Source. GCIA Gold Certification (2008)
22. KnowEm LLC. Social media brand search engine. http://knowem.com/
23. Krombholz, K., Hobel, H., Huber, M., Weippl, E.: Social engineering attacks on the knowledge worker. In: Proceedings of Security of Information and Networks, SIN 2013, pp. 28–35. ACM (2013)
24. LinkedIn. Business social networking service. http://linkedin.com/
25. MarketVisual. Business search engine. http://www.marketvisual.com/
26. Milosevic, N.: Introduction to Social Engineering (2013)
27. Mitnick, K.D., Simon, W.L.: The Art of Deception: Controlling the Human Element in Security (2003)
28. Monster Wolrdwide Inc., Job search engine. http://monster.com/
29. Namechk. Username and domain search tool. https://namechk.com/
30. National Association of Counties. http://www.naco.org/
31. Pakistan Government. Federal board of revenue. http://www.fbr.gov.pk/
32. Paterva. Maltego clients and servers. https://www.paterva.com/web6/products/maltego.php

33. Public Accountability Initiative. http://littlesis.org/
34. Ratsit & Invativa. Credit business website. http://www.ratsit.se/
35. Regan, K.: 10 Amazing Social Media Growth Stats From 2015 (2015)
36. Shodan. Search engine for the internet of things. https://www.shodan.io/
37. Socialmention. Social media search platform. http://socialmention.com/
38. Spokeo. People search website. http://www.spokeo.com/
39. The Internet Archive. The wayback machine. https://archive.org/web/
40. Tomes, T.: Web reconnaissance framework. https://bitbucket.org/LaNMaSteR53/recon-ng
41. Verizon. Data Breach Investigations Report (2012). http://www.verizonenterprise.com/resources/reports/rp_data-breach-investigations-report-2012-ebk_en_xg.pdf
42. Verizon. Data Breach Investigations Report (2013). http://www.verizonenterprise.com/resources/reports/rp_data-breach-investigations-report-2013_en_xg.pdf
43. Warkentin, M., Willison, R.: Behavioral and policy issues in information systems security: the insider threat. Eur. J. Inf. Syst. **18**(2), 101–105 (2009)
44. Watson, G., Mason, A., Ackroyd, R.: Social Engineering Penetration Testing: Executing Social Engineering Pen Tests, Assessments and Defense. Syngress, Rockland (2011)
45. Xing. Business social networking service. http://xing.com/

Erratum to: A Structured Comparison of Social Engineering Intelligence Gathering Tools

Kristian Beckers[1(✉)], Daniel Schosser[1], Sebastian Pape[2], and Peter Schaab[1]

[1] Institute of Informatics, Technische Universität München (TUM), Boltzmannstr. 3, 85748 Garching, Germany
beckersk@in.tum.de

[2] Faculty of Economics and Business Administration, Goethe University Frankfurt, Theodor-W.-Adorno-Platz 4, 60323 Frankfurt, Germany

Erratum to:
Chapter "A Structured Comparison of Social Engineering Intelligence Gathering Tools" in: J. Lopez et al. (Eds.): Trust, Privacy and Security in Digital Business, LNCS 10442, https://doi.org/10.1007/978-3-319-64483-7_15

The presentation of Table 7 was incorrect in the original version of this chapter. The correct version is given below:

Table 7. Tools vs. AttackType knowledge with P for phishing, I for impersonation, and B for baiting

	Cree.py	Gitrob	KnowEm	LinkedIn	Maltego	Namechk	Recon-ng	Spokeo	theHarvester	Wayback Machine	Wireshark	Xing
Telephone Number							P					P
Friends			P,I	P,I	P,I	P,I						P,I
Personal Information	P,I		P,I	P,I		P,I		P,I				P,I
Private Locations	P,I							P,I				P,I
E-Mail				P	P		P		P			P
InstantMessenger			P		P	P		P				P
Co-Workers: NewEmployee				I	I							I
Co-Workers: Hierarchies				I			I					I
Lingo												
Facilities: Security-Measures		B,I									B,I	
Facilities: Company Location	B,I			B,I			B,I	B,I	B,I			B,I
Websites					P		P		P	P		

The original chapter has been corrected.

The updated online version of this chapter can be found at https://doi.org/10.1007/978-3-319-64483-7_15

J. Lopez et al. (Eds.): TrustBus 2017, LNCS 10442, p. E1, 2017.
https://doi.org/10.1007/978-3-319-64483-7_16

A.6 PERSUADED: Fighting Social Engineering Attacks with a Serious Game

Dina Aladawy, Kristian Beckers, and Sebastian Pape. PERSUADED: Fighting Social Engineering Attacks with a Serious Game. In Steven Furnell, Haralambos Mouratidis, and Günther Pernul, editors, *Trust, Privacy and Security in Digital Business - 15th International Conference, TrustBus 2018, Regensburg, Germany, September 5-6, 2018, Proceedings*, volume 11033 of *Lecture Notes in Computer Science*. Springer, 2018. ISBN 978-3-319-98384-4. doi: 10.1007/978-3-319-98385-1_8. URL https://doi.org/10.1007/978-3-319-98385-1_8

PERSUADED: Fighting Social Engineering Attacks with a Serious Game

Dina Aladawy[1], Kristian Beckers[1,3], and Sebastian Pape[2,3(✉)]

[1] Institute of Informatics, Technische Universität München (TUM), Boltzmannstr. 3, 85748 Garching, Germany
[2] Faculty of Economics and Business Administration, Goethe University Frankfurt, Theodor-W.-Adorno-Platz 4, 60323 Frankfurt am Main, Germany
sebastian.pape@m-chair.de
[3] Social Engineering Academy (SEA) GmbH, Eschersheimer Landstrasse 42, 60322 Frankfurt am Main, Germany

Abstract. Social engineering is the clever manipulation of the human element to acquire information assets. While technical security of most critical systems is high, the systems remain vulnerable to attacks from social engineers. The challenge in defeating social engineering is that it is a deceptive process that exploits human beings. Methods employed in social engineering do not differ much from those used to perform traditional fraud. This implies the applicability of defense mechanisms against the latter to the context of social engineering. Taking this problem into consideration, we designed a serious game that trains people against social engineering using defense mechanisms of social psychology. The results of our empirical evaluation of the game indicate that the game is able to raise awareness for social engineering in an entertaining way.

Keywords: Security controls · Social psychology · Gamification

1 Introduction

Chris Hadnagy [9] defines social engineering as "Any act that influences a person to take an action that may or may not be in their best interest". Kevin Mitnick told in an interview the following about the relevance of social engineering: "The hacker is going to look at the weakest link in the security chain, [...] if they see it's your people – if you don't educate your people about social engineering and they're easy targets – then that's where the attacker is going to attack." [6] Mitnick's statement was made over a decade ago and is still of utmost importance today as several current studies confirm [4,15].

In a previous work, we provided a mapping between social psychology and IT-security regarding Social Engineering defence [17]. In particular, we analysed social psychology methods of training against persuasion and mapped them to trainings in IT security. One identified gap is the lack of using *inoculation*, the repeated confrontation of people with a challenging situation in order to trigger

S. Furnell et al. (Eds.): TrustBus 2018, LNCS 11033, pp. 103–118, 2018.
https://doi.org/10.1007/978-3-319-98385-1_8

an appropriate response. Our contribution in this work is filling the identified gap with a serious game called *Persuaded*.

Djaouti et al. [5] define serious games as follows "A serious game or applied game is a game designed for a primary purpose other than pure entertainment.". We choose a serious game, because games recently built a reputation for getting employees of companies involved in security activities in an enjoyable and sustainable way. Williams et al. [20] introduced the protection poker game to prioritise risks in software engineering projects. Shostack [18] from Microsoft presented his Elevation of Privileges card game to practice threat analysis with software engineers. Furthermore, games are used as part of security awareness campaigns [7] and particularly as a part of social engineering threat analysis [1].

Our contribution Persuaded has inoculation incorporated into the core game mechanics to trigger resistance to social engineering attacks through exposing people to realistic attack scenarios. We designed our serious game to achieve the following goals: (a) increasing awareness of social engineering, (b) training resistance to persuasion and (c) addressing the general population. In order to provide the validity of the attack scenarios, we took all of them from scientific publications. The game enables employees to learn about social engineering, while practicing simultaneously. This immediate application of learned knowledge has proven to have lasting effects [8].

The game works as follows. Employees get confronted with a possible social engineering threat and have to select a defense mechanism. This defense mechanism is a pattern of behaviour ensuring a secure outcome. For example, an employee gets a phishing mail and is asked to open its attachment. Afterwards the player selects a countermeasure: "Do not open the email and inform the information security department immediately". The player gets immediate feedback whether the chosen defense is correct. In particular, the offered defenses can be part from a company's security policy. Non surprisingly, Soomro et al. found that development and execution of information security policy had a significant impact on the quality of management of information security [19]. Earlier, Pahnila et al. already concluded that appraisal and facilitating conditions have significant impact on attitude towards complying with the security policy while sanctions and awards do not have a significant effect on the intention to comply [14]. Thus, enabling employees to become familiar with the security policy in a playful way contributes to the holistic security of the company.

The remainder of the paper is organised as follows: We start with an overview of related work (Sect. 2) and a description of our game (Sect. 3). In the next sections we describe the study and its results. We end with a discussion of the results, threats to validity and the conclusion.

2 Related Work

As security is usually a secondary task, computer security training has often been perceived to be an uninteresting enforcement to users and managers. The approach of developing serious games has therefore been adopted to provide knowledge and training in that field.

CyberCIEGE is a role playing video game, where a player acts as an information security decision maker in an enterprise. Players' main responsibilities are to minimize the risk to the enterprise while allowing users to accomplish their goals. Similar to Persuaded, the game offers a simulation of the reality particularly portraying the need to maintain the balance between productivity and security. As decision makers, players get to make choices concerning users (i.e. How extensive will background checks be?), computers (i.e. How will computers be networked?) and physical security (i.e. Who is allowed to enter a zone?) while monitoring the consequences of their choices. When compared to Persuaded, we recognized CyberCIEGE offered several advantages common to those offered by Persuaded. For instance, players are in a defensive mode and they get to make decisions and experience their consequences. CyberCIEGE even incorporates assets and resources in the game, which is a missing element in Persuaded. On the other hand, the game requires longer time to learn and to play [10].

PlayingSafe is a serious game in the domain of social engineering. It consists of multiple choice questions which are wrapped in typical mechanics of a board game. Since questions provided are exclusive to social engineering, the game is very similar to ours. The main difference lies however in the focus in the topic of social engineering. PlayingSafe asks questions in the fields of Phishing, advanced fee fraud, spam and others, being a category that covers less common attacks. Our game on the other hand covers a broader field without offering depth in each topic. Additionally, our game incorporates strategy favouring the entertainment element, in order to enhance the game experience the game provides [12].

SEAG is a serious game designed to raise awareness of social engineering. The game utilizes levels that tackle different cognitive aspects and hence provide an effective learning experience. The first level consists of quiz-like questions to build a knowledge base for the players. The second level is a match game where players have to match social engineering terms with respective pictures. Finally, the players are presented with real life scenarios to analyse pertaining to threat. This simulation of real life application of the learnt lesson should test players ability to detect attacks- an approach very similar to inoculation [13].

3 Game Description

To fill the gap, identified by Schaab et al. [17], we designed a game that does not only provide knowledge, but rather trains people by implementing theories from social psychology on the resistance to persuasion. In this section, we give a brief overview of key design decisions, their rationale and our goals (cf. Fig. 1).

Game Requirements: We refined our goals and report them in the following categorised by key areas of game design.

Ease to Learn: A low level of complexity allows to learn the game more easily, and thus is more attractive to novices in game play.

Ease to Play: To be easily integrated into the players' daily routine, the game should have a minimum of necessary preparations and a short play time. Given online games require less preparation than tabletop games, it should be online.

106 D. Aladawy et al.

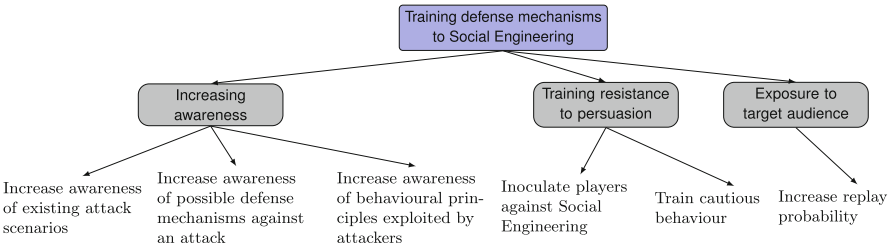


Fig. 1. Definition of goals for the game

Replay Value: The replay value depends to a large amount on the ease to learn and play of the game. In order to maintain the appeal to expert players as well, game mechanics should provide a substantial entertainment element along with long term motivation and challenging the players. As multi-player games depend on the availability of other players, a single player game is preferred.

Player's Role: In order to inoculate players against social engineering, they have to be in the position of an attack receiver.

Textual Content: Since our awareness goals cater for presenting attack/defense scenarios, the game design should support the presentation of textual content.

Game Mechanics: In order to create a single player game with easy rules and low complexity, we decided to aim for a patience and solitaire game approach [11] instead of e.g. involving machine learning approaches [3] which would tend to result in a game with multi-player feeling. Thus, the player may choose between playing cards from his/her hand or draw the next card from the deck. As known from patience games, the deck is shuffled automatically for each game.

Types of cards and card functionalities: Four types of cards were chosen.

1. *Attack* cards include attack scenarios in textual form.
2. *Defense* cards describe a pattern of behaviour that protects the player against an exploitation attempt. A defense card exists for each attack card.
3. *See The Future* cards allow the player to take a peek on the three upper cards in the card deck.
4. *Skip turn* cards allow the player to take the upper card of the deck and put it below the deck.

Mechanics and rules: A turn in Persuaded consists of the following rules:

1. Play an action card or draw a card from the deck.
2. If you draw any card that is NOT an Attack, the turn is over. Put the card to your hand cards.
3. If you draw an Attack card, you *have to* play a Defense card. The correct (wrong) defense gains you 10 (−5) points. The Defense card is only discarded if you had a correct match. Otherwise it's put back in the deck.

4. If you draw an Attack card and don't have *any* Defense card in your hand, you lose one heart (life). If you lost all three hearts the game is over.
5. The game is won if the deck is empty and is lost if the player loses all 3 lives before finishing the deck.

These mechanics have several consequences. Drawing an Attack card *forces* the player to play a Defense card. Thus, even if a player notices he has no matching defense, he has to burn a defense card. This was introduced to further encourage cautiousness when drawing cards from the deck. The player needs to use *See the future* cards to have a peek on the pile and then postpone attacks if he does not have a matching defense by playing a *Skip turn* card. This also forces the player to match upcoming attacks and defenses in hand before drawing form the pile.

Long Term Motivation: As known from patience games, the deck is shuffled automatically for each game. This causes each game to be different from the game(s) before. Thus, the player needs to come up with different moves to win the game and can not simply try until he/she finds the 'optimal solution'. Additionally, the introduced randomness, causes Attack cards to appear before their respective Defense cards in the deck. Therefore – if action cards are not distributed accordingly – this may lead to situations where the player simply has to guess what might be the 'best next move'. The idea behind this rationale is that not only has the player to learn how to make best use of "See the future"- and "Skip turn"-cards, but also needs to have some luck in order to achieve the best possible score. We balanced it in a way, that it is always possible to win, but might not be possible to get the maximum score.

Game Content: In order to provide the knowledge needed to increase players' awareness, scenarios of attacks and their respective defenses were incorporated in the game. We selected eight attack scenarios that represent different social engineering attack types, namely Baiting, Phishing, Tailgating, Mail attachment, physical and virtual Impersonation, Voice of Authority and Popup Windows. The attacks were inspired by a card game for eliciting security requirements [2]. Defense cards, on the other hand, confronted us with challenges, as it is not very intuitive to act against behavioural principles, which is exactly the element exploited by social engineering. We identified explicit defenses encouraged from best practice by security departments in different companies. Initially, defenses were meant to be generic and applicable for several attack scenarios. However, resulting from our selection of proposed scenarios, we noticed, that all had similar generic defenses, i.e. to verify the source or the person. Hence, we decided to incorporate one-to-one matches thereby providing eight specific Defense cards.

Game Interface Design: In confirmation with Don Normann's Design principles [16] for user interface design, we opted for an intuitive user interface that adheres to the needs of novices as well as experts in game play. The proposed design was further tested and adapted according to the feedback we received during the piloting phase.We used different colors for each type of the *cards* (see Fig. 2). For the attack/defense scenarios, we kept the text as short as possible and divided the content in up to three bullet points. Action cards consist of graphics

Fig. 2. From left to right: Attack card, Defense card, Skip card, See the future card

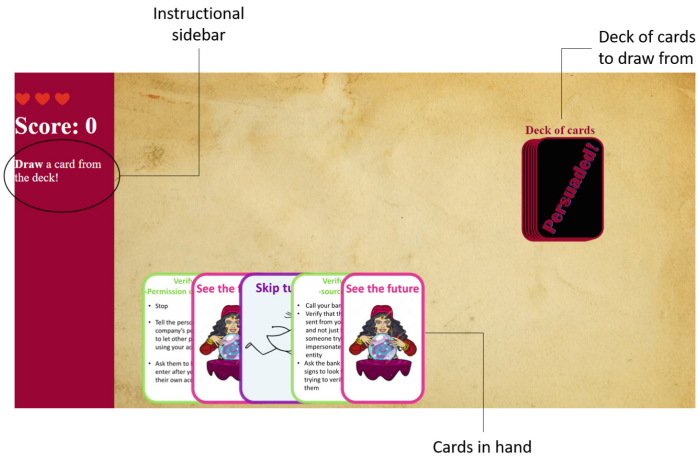


Fig. 3. Initial game setting

that reflect their functionality, attack and defense cards have titles summarizing their content. However, titles of matching pairs are not the same. This design decision was intentionally incorporated, in order to assure that players have to read the cards' contents. The *Game Setting* (see Fig. 3) was designed to be both intuitive and informative.

Cards in Hand: The overlapping display of the current cards in hand simulates the holding of cards in real life (cf. Normann's mapping principle). When a player moves the mouse over a card, this card is emphasized by moving the other cards to the left and right to allow the player a complete view of the card. This enhances the player's experience while maintaining readability of the content.

Scoring: As score and lives function as a reward and punishment system, it is important to make sure, they capture players' attention when they change. Therefore, we decided to reflect modifications of scores and lives using dynamic

feedback. In addition to using coloured terms such as "Defended", "Wrong match" and "Persuaded", the decrease or increase of score and lives is at the top left.

We show the game in detail in our Video Tutorial for Persuaded[1]. Furthermore, we stored the data of our experiment and an extensive technical report online[2].

4 Study Design

Prior to conducting the case study, Persuaded has been evaluated through several rounds during the design and the development phase. First of all, the scenarios were tested for suitability of attacks and defenses, in addition to the ease of understanding of the presented content. Following this, the game's functionality and mechanics were tested during a piloting phase. The participants for the pilot tests were very heterogeneous. Play tests and semi-structured interviews were conducted with 3 security experts, a psychology expert, a games engineering student, an informatics student and a philosophy student. Feedback provided in this phase, was largely incorporated in the design and the implementation. The content was reviewed by 2 security experts and 2 informatics students. The scenarios' text was reviewed by a student in Translation Studies.

4.1 Preparation and Collection of Data

The flow of a session with a subject consisted of the following steps:

1. Answer the pre-questionnaire.
2. Watch the game tutorial as many times as you need.
3. Ask questions about the game rules.
4. Play the game.
5. Answer the post-questionnaire.

We employed first and second degree methods for our data collection. Before the session started, subjects were encouraged to provide feedback throughout the session. Many subjects took this into account and offered valuable feedback on the questionnaire, the game and the tutorial. Some subjects even played the game in a think aloud mode, which turned out to be very useful feedback. Furthermore, second degree data was collected during the game play to evaluate to what extent the game adheres to requirements specified in prior sections. We logged all decisions made during the game, making it possible to replicate the entire round. In addition, the time to play as well as the final score and number of lives left was collected. This enabled us to analyse the effects of our random factor on the entire game experience.

[1] https://youtu.be/UWhc1e6ngd0.
[2] https://sites.google.com/site/researchpersuaded/.

Pre- and Post Questionnaires. The effect of inoculation can be measured by observing peoples' reactions to stronger persuasive attacks as the ones they were inoculated with. This implied that we have to present players with stronger scenarios of social engineering after the game in order to be able to derive whether it was effective or not. This however, was not enough as an effect measurement as we were not aware, whether people were vulnerable before the game at all or not. Hence, we decided to conduct questionnaires before and after the game was played. The questionnaires presented social engineering scenarios as single choice questions, where players had to choose one of the given behaviours as a reaction to the given situation. The same scenarios with the same reactions were presented both in the pre-and post-questionnaire. This was intentionally done in order to be able to measure effects of the game as change of answers. In addition to the situations presented in the pre-questionnaire, demographic data was collected to draw conclusions for different types of people given our exposure goals. Moreover, data concerning technical background was collected which might be relevant to scenarios such as *Phishing* and *Popup Window* as well as malicious *Mail Attachments*. Lastly, items were used to measure background knowledge of social engineering and to measure the subjective perception of vulnerability in order to have an indicator of optimism bias. Players were also asked to indicate at which point they understood the game to measure the learning curve and the effectiveness of the tutorial and whether they would play this game again or not.

4.2 Data

The equation introduced to evaluate the questionnaires was:

Learning outcome = $\sum$ security-aware behaviour in post-questionnaire − $\sum$ security-aware behaviour in pre-questionnaire.

For quantitatively evaluating the players' decisions throughout the game, we relied on the following data.

Matching of Attacks and Defenses. We used a half automatic analysis process to measure the number of attacks that were correctly defended as well as the number of burned defense cards. This data maps the understandability of the content of the cards. Moreover, as we are not game designers we decided to use them as an indicator of the impact of certain game elements such as the randomness of the cards' order and the variability of Attack and Defense cards.

Usage of Action Cards. We also collected data concerning the number of cards that were foreseen and the number of cards that were unknowingly drawn. This information was not only used to evaluate the game flow, given that Flow cards are key elements of the winning strategy. They were also used as an indicator of players' risk behaviour and alertness during the game play.

Reward and Punishment System. Lastly we collected the score data as well as the number of lives left. This information was employed to test whether our reward and punishment systems are effective or whether they are influenced by the random element in the game.

5 Results

The study was conducted with 21 participants including 9 female and 12 male participants. The age ranged from 19 to 35 years. Given our exposure goals, we sampled subjects with different backgrounds regarding their studies. 16 of the participants indicated they are university students, while 5 are currently pursuing an academic career. We disregard one participant's results. These were invalid due to changes of the content of the questionnaires. In contrast to the variation in age and occupation, our sample is very homogeneous in technical background. It is important to mention that at this point, we only consider technical background in relation to how often the computer is used, which is sufficient for understanding the game content. Answering this question, 95% indicated they use their computers daily while 43% use it daily for job related matters.

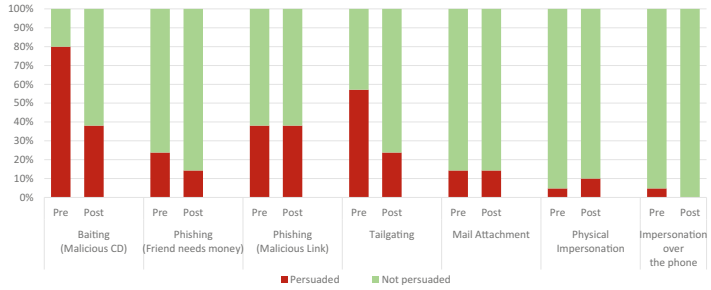


Fig. 4. Number of reactions reflecting falling for the attack in comparison to security aware reactions before and after the game

5.1 Results Relevant to Inoculation

Our game is an implementation of inoculation against social engineering. Its effectiveness as a training method was evaluated using pre- and post-questionnaires in addition to several metrics.

Reactions to Situations. Participants were given social engineering scenarios and asked to choose one reaction, they were most likely to adopt if confronted with such a scenario. The questions proposed three answers with one mapping the security-aware behaviour when encountering a potential threat and the other two options reflecting extreme reactions. The first extreme is a paranoid reaction, whereas the second reflects falling victim for the attack. Results from the pre-questionnaire show that in 5 of the 7 scenarios the majority of the participants would have behaved in a manner that would not endanger them. In the other two scenarios, a high number of subjects would have fallen for the attack.

The results of the post-questionnaire show significant differences. For the *Tailgating* scenario which describes the situation of meeting a strange lady who

is locked out of the house building and whether a person should verify her identity before letting her in or not, the number of participants indicating they would behave in a security-aware manner rises from 43% to 76%. Nevertheless, the *Baiting* attack which questions whether free handed CDs from street musicians should be scanned or not, remains the one scenario where the reaction indicating falling for the attack is the one chosen the most. For the *Phishing (Malicious Link)* and *Mail attachment* attack, the numbers do not show significant change. For the remaining scenarios only slight changes are noticeable, once even favouring the rise of number of participants who would fall for the attack as it is the case in the *Physical impersonation* attack. Figure 4 shows an overview of the change in responses triggered by our game. Given inoculation relies on repeatedly confronting individuals with mild persuasive attacks, we also measured the number of times players read an attack card in the game which indicates that each attack is read 1.5 times in average.

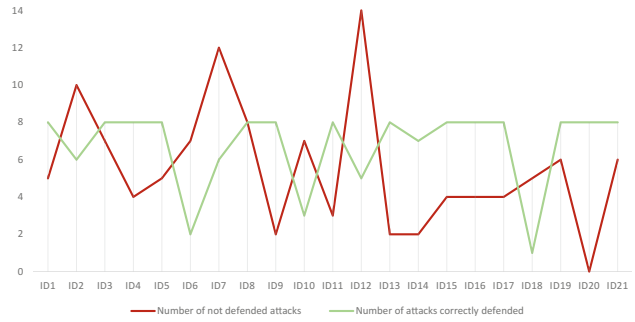


Fig. 5. Defended vs. Not defended attacks for all participants

5.2 Reward/Punishment System

The maximally achievable score is 80 points if the player did not make any wrong match or if the player did a wrong match at the beginning of the game when the score was still 0. Only one player was able to score 80 points and 2 players could score 75 with an average score of 51 points. The majority of players achieved a score of 65 points. Considering the lives maintained in the game, 15 players were able to finish the deck maintaining at least one heart while 6 others lost the game before finishing the deck due to losing their lives (Fig. 5).

5.3 Time to Play

The time needed to play the game ranged from 02:53 min to 16:03 with an average of 08:09 min. We further differentiate the time to play needed to win the game by finishing the cards in deck and the time to play for lost games. The range measured for games that were won through finishing the deck lies between 05:05 and 16:03 min with an average of 8:33 min.

5.4 Matching of Cards

For 15 of the participants, the number of successfully defended attacks is higher or equal to the number of not defended ones. The latter further includes attacks that were drawn without having defense cards in hand. Not defended attacks can be further categorised in mismatched attacks (75% of not defended attacks) and attacks that were drawn without having defense cards in the hand (25% of not defended attacks).

We further distinguish mismatched attacks in those, where the player had the matching defense in hand, as in the player is accountable for the mismatch and those, where the player was forced to play a Defense card. For all participants the number of burned defenses is higher than the number of *truly* mismatched attacks. Moreover, 66% of the participants did not even once mismatch an attack while the matching Defense card is in hand.

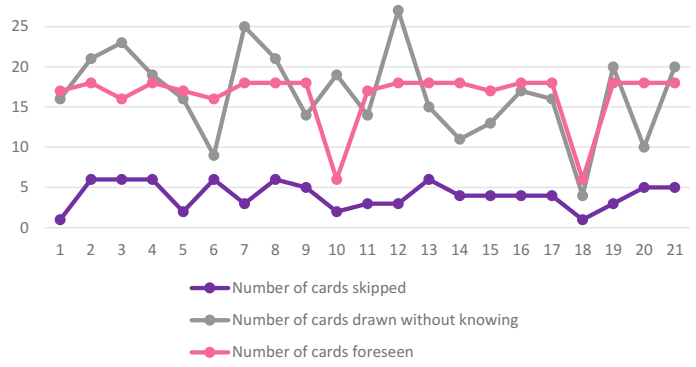


Fig. 6. Overview of skipped cards, cards foreseen and cards drawn blindly

5.5 Action Cards

The key strategy to winning the game is to use *See The Future* cards and then avoid attacks, whose defenses are not in hand, with *Skip* cards. This is why we analysed players' usage of action cards considering players can see a maximum of 18 cards before drawing them and skip a maximum of 6 cards (Fig. 6).

We further analysed players risk behaviour according to when cards are drawn blindly despite having a *See The Future* card. Our results show that a total of 16 players have drawn at least one card without seeing it, having a *See The Future* card in the hand. An average of 2.6 cards were drawn blindly despite having the chance to foresee them. More importantly, however, is the number of cards drawn blindly despite having both a *See The Future* card and a *Skip* card. This card combination would have offered the chance to knowingly avoid drawing that card. This was done by 10 of the participants with an average of 1.6 cards drawn blindly despite having the chance to knowingly avoid them.

5.6 Learning Curve and Replay Value

Finally, subjects were asked to indicate whether they would play this game again. 17 participants (81%) expressed that they would play this game again while the remaining 4 participants claimed they would not. When asked about the understandability of the game mechanics, 14 people (67%) mentioned they understood the game right away (following only the tutorial) while the remaining 7 participants needed some turns to fully understand how the game works.

6 Discussion

Through the conduction of interviews with the players, we could collect feedback that is of value to future work. More importantly, the feedback showed potential threats to the validity of the data collected in the questionnaires.

6.1 Feedback on Pre-Questionnaire:

The pre-questionnaire included social engineering scenarios, where players had to chose a reaction. Particularly, the baiting scenario, where street musicians would intentionally offer malicious CDs, was perceived by two participants to be an "interesting new attack, [they] have never thought of". The tailgating attack was stated to be relevant to one of the participants. Particularly these two attacks were the only ones in the pre-questionnaire, where most participants chose the reaction, that would favour the attackers intentions. We conclude, that these attacks were new to the participants choosing that reaction. This is backed up by the interview comments in addition to the results of the final question in the post-questionnaire where seven participants indicated the *Tailgating* attack was new to them while four indicated the same for the *Baiting* attack. Improvement suggestions were to incorporate an "other" option as a possible reaction to the situations and to collect data on the used operating system, given it implies a certain security level provided by the technology alone.

6.2 Feedback on the Game Mechanics:

We received extensive feedback on our randomness factor of our game.

General Perception: The general perceptions during game play provided feedback that conforms with our design goals. One participant stated, that for them the game simulates the reality. The player further explained, that in real life, it is rather difficult to expect social engineering attacks and always be ready for them, which they found was mapped through the random factor. Furthermore, the player mentioned that usually even the most cautious people might fall victim for social engineering again supporting the vulnerability in the game, where players are not able to defend themselves when drawing attacks before their respective defenses. Another participant provided feedback on the challenge level in the game saying that "one has to think". For this player, the game

was also "easy to understand", reflecting the modesty of the trade-off between those two conflicting elements. Finally, the player emphasised the importance of the game being single player for the replay value, saying that he can "play the game another 3 times just right now". This data conforms with the data collected on replay in the post-questionnaire underlining the high potential for replay of the game.

Understanding of the Game Mechanics: We opted for ease of learning, realised by simple mechanics, a detailed tutorial and an intuitive game interface design. Several questions asked during game play, however, indicate otherwise. Misconceptions and uncertainty were particularly common regarding the functionality and usage of action cards. Examples for questions, we received concerning action cards are: *What does a "See the future" card really do?*, *What does a "Skip" card really do?*, *How many cards are skipped by playing a "Skip" card?*, *Will skipped cards be added in the deck?*. We cannot determine, whether these questions were asked due to lack of understanding or rather to confirm prior understanding of the functionality. However this data explains the relatively small numbers of wasted *See The Future* cards and *Skip* cards, which were played without having foreseen what was being skipped. We assume the wrong usage of the action cards happened at the beginning of the game, as four players have indicated, they needed some turns to fully grasp the game mechanics. Five players asked for the number of cards in the deck. We assume, this was asked in order to develop certain strategies rather than to indicated extensive length of the game duration, which is further supported by our measurement of game duration being 09:45 min in average.

Card Content: The serious element of Persuaded lies in the content of the cards. This is why, it is important to monitor whether cards are read in detail or not. Four players indicated after the first couple turns, that they have not been reading the cards' contents, while two others attempted to match the titles of attack and defense cards in the beginning. Still, all six players started reading the content of the cards after a couple of turns. We assume this was motivated by the punishments they received for wrong matches. This data is particularly relevant to data collected on the number of mismatched cards while having the correct defense in hand. Given this only happened to an average of 0.52 cards, we build the assumption that *truly* mismatched cards were rather a result of not reading the cards than a result of the complexity of the content. Still, one player further indicated, that the match between attacks and defenses was not always clear. This was however intentionally incorporated in the design, as we wanted players to reflect about the scenarios and the defenses instead of recognizing the matches from the cards. Furthermore, one player suggested, it would be better to see all the cards before playing the game to create a mental scheme of matching cards. Thereby, players can solely focus on training the strategy during the game and would strengthen the mental scheme by recalling matches between cards.

Randomness in the Game: The randomness of the game was not very welcomed by the participants. Although one player indicated, it provided a

simulation of real life, four other players perceived themselves to have no control in the game with two players evaluating the game as unfair. An important aspect influenced by the randomness is replay value. Replay value is usually supported by the probability of the player to excel in the game play. Having a random factor largely limits improvements in the game as players' decisions are only partially relevant for the game results. This was further confirmed by two players, who said they would only play the game again, if they could get better at it.

7 Threats to Validity

We discuss potential threats to the validity according to Wohlin [21].

Construct Validity. Questions in the post-questionnaire are supposed to indicate probable reactions of the participants to given scenarios. There is, however, a possibility that participants remember their answers to the same questions during the pre- questionnaire. However, if players are aware, the game has educational purpose, this might lead to a conscious choice of the correct answers to indicate having understood the content. In addition, several metrics were derived from players' decisions during the game session as explained in the previous section. This data is however subject to effects of concentration and motivation during the game. Moreover the results assume that the functionality of the game and the different cards is understood at the beginning of the game in contrast to the feedback received on the learning curve.

Internal Validity. We measure the learning outcome as the difference between the sum of correctly answered questions in the pre- and post-questionnaire. We cannot determine whether players are inoculated by the scenarios of the game or by the scenarios mentioned in the pre-questionnaire as these also reveal persuasive arguments used by social engineers. This effect of an inoculation at an early point is attempted to be overcome by hiding the subject of the study from participants until the questions of the pre-questionnaire are answered.

External Validity. We conducted the case study with a heterogeneous population regarding their educational background and could identify acceptability of the game even for subjects without prior knowledge in security or social engineering. However, our results regarding the effectiveness and the learning outcome of the game are to be considered taking the random factor of the game and other threats to validity into account.

8 Conclusion

We designed, implemented and evaluated a serious game for training social engineering defense mechanisms, called "Persuaded". Several goals were specified and refined to achieve the serious purpose of the game: *Increase awareness:* of attack scenarios, defense mechanisms and exploited behavioural principles. *Train resistance to persuasion* by inoculation against social engineering and to train cautious behaviour. Finally, to cater for *exposure to the general population* through

increasing replay probability and ease of understanding of the social engineering threat. Results of our case study indicate great potential for the application of social psychology defense mechanisms to social engineering. Our serious game offers a tool for monitoring decision making processes and risk-taking behaviour. More importantly, it was successful at raising awareness to new attack scenarios in an entertaining way such that people would enjoy learning about social engineering and how they can defend themselves against it.

Acknowledgements. This research has been partially supported by the Federal Ministry of Education and Research Germany (BMBF) with project grant number 16KIS0240.

References

1. Beckers, K., Pape, S.: A serious game for eliciting social engineering security requirements. In: Proceedings of the 24th IEEE International Conference on Requirements Engineering, RE 2016, pp. 16–25. IEEE Computer Society (2016)
2. Beckers, K., Pape, S., Fries, V.: HATCH: hack and trick capricious humans - a serious game on social engineering. In: Proceedings of British HCI 2016, pp. 1–3. ACM (2016)
3. Bowling, M., Fürnkranz, J., Graepel, T., Musick, R.: Machine learning and games. Mach. Learn. **63**(3), 211–215 (2006)
4. Dimensional Research: The Risk of Social Engineering on Information Security: A Survey of IT Profesionals (2011). http://docplayer.net/11092603-The-risk-of-social-engineering-on-information-security.html
5. Djaouti, D., Alvarez, J., Jessel, J.-P.: Classifying serious games: the G/P/S model. In: Handbook of Research on Improving Learning and Motivation Through Educational Games: Multidisciplinary Approaches, pp. 118–136 (2011)
6. ENISA: Social engineering: exploiting the weakest links. Whitepaper, October 2008. https://www.enisa.europa.eu/publications/archive/social-engineering
7. Gondree, M., Peterson, Z.N.J., Denning, T.: Security through play. IEEE Secur. Priv. **11**(3), 64–67 (2013)
8. Greitzer, F.L., Kuchar, O.A., Huston, K.: Cognitive science implications for enhancing training effectiveness in a serious gaming context. J. Educ. Resour. Comput., **7**(3), (2007)
9. Hadnagy, C.: Social Engineering: The Art of Human Hacking. Wiley, Hoboken (2010)
10. Irvine, C.E., Thompson, M.F., Allen, K.: Cyberciege: gaming for information assurance. IEEE Secur. Priv. **3**(3), 61–64 (2005)
11. Morehead, A.H.: The Complete Book of Solitaire and Patience Games. Read Books Ltd., Redditch (2014)
12. Newbould, M., Furnell, S.: Playing safe: a prototype game for raising awareness of social engineering. In: Australian Information Security Management Conference, p. 4 (2009)
13. Olanrewaju, A.-S.T., Zakaria, N.H.: Social engineering awareness game (SEAG): an empirical evaluation of using game towards improving information security awareness. In: Proceedings of the 5th International Conference on Computing and Informatics, ICOCI 2015 (2015). Accessed 16 Oct 2016

14. Pahnila, S., Siponen, M., Mahmood, A.: Employees' behavior towards IS security policy compliance. In: 40th Annual Hawaii International Conference on System Sciences, HICSS 2007, p. 156b. IEEE (2007)
15. PWC: Information Security Breaches Survey (2016). https://www.pwc.be/en/documents/media-centre/publications/2016/information-security-breaches-survey-2016.pdf
16. Rogers, Y., Sharp, H., Preece, J., Tepper, M.: Interaction design: beyond human-computer interaction. netWorker: Craft Netw. Comput. **11**(4), 34 (2007)
17. Schaab, P., Beckers, K., Pape, S.: Social engineering defence mechanisms and counteracting training strategies. Inf. Comput. Secur. **25**(2), 206–222 (2017)
18. Shostack, A.: Threat Modeling: Designing for Security, 1st edn. Wiley, Hoboken (2014)
19. Soomro, Z.A., Shah, M.H., Ahmed, J.: Information security management needs more holistic approach: a literature review. Int. J. Inf. Manage. **36**(2), 215–225 (2016)
20. Williams, L., Meneely, A., Shipley, G.: Protection Poker: the new software security "game". IEEE Secur. Priv. **8**(3), 14–20 (2010)
21. Wohlin, C., et al.: Experimentation in Software Engineering: An Introduction. The Kluwer International Series in Software Engineering. Springer, Boston (2012). https://doi.org/10.1007/978-1-4615-4625-2

A.7 PROTECT - An Easy Configurable Serious Game to Train Employees Against Social Engineering Attacks

Ludger Goeke, Alejandro Quintanar, Kristian Beckers, and Sebastian Pape. PROTECT - an easy configurable serious game to train employees against social engineering attacks. In *Computer Security - ESORICS 2019 International Workshops, IOSec, MSTEC, and FINSEC, Luxembourg City, Luxembourg, September 26-27, 2019, Revised Selected Papers*, volume 11981 of *Lecture Notes in Computer Science*, pages 156–171, Cham, 2019. Springer International Publishing. ISBN 978-3-030-42051-2. doi: 10.1007/978-3-030-42051-2_11. URL https://link.springer.com/chapter/10.1007/978-3-030-42051-2_11

PROTECT – An Easy Configurable Serious Game to Train Employees Against Social Engineering Attacks

Ludger Goeke[1], Alejandro Quintanar[1], Kristian Beckers[1], and Sebastian Pape[1,2(✉)]

[1] Social Engineering Academy (SEA) GmbH, Eschersheimer Landstrasse 42, 60322 Frankfurt am Main, Germany
[2] Faculty of Economics and Business Administration, Goethe University Frankfurt, Theodor-W.-Adorno-Platz 4, 60323 Frankfurt am Main, Germany
sebastian.pape@m-chair.de

Abstract. Social engineering is the clever manipulation of human trust. While most security protection focuses on technical aspects, organisations remain vulnerable to social engineers. Approaches employed in social engineering do not differ significantly from the ones used in common fraud. This implies defence mechanisms against the fraud are useful to prevent social engineering, as well. We tackle this problem using and enhancing an existing online serious game to train employees to use defence mechanisms of social psychology. The game has shown promising tendencies towards raising awareness for social engineering in an entertaining way. Training is highly effective when it is adapted to the players context. Our contribution focuses on enhancing the game with highly configurable game settings and content to allow the adaption to the player's context as well as the integration into training platforms. We discuss the resulting game with practitioners in the field of security awareness to gather some qualitative feedback.

Keywords: Security controls · Social psychology · Serious games · Fraud prevention · Security training

1 Introduction

Kevin Mitnick a most famous social engineer was interviewed over 15 years ago and stated the following. "The biggest threat to the security of a company is not a computer virus, an unpatched hole in a key program or a badly installed firewall. In fact, the biggest threat could be you [...] What I found personally to be true was that it's easier to manipulate people rather than technology [...] Most of the time organizations overlook that human element" [3]. Today this is as true as it was back than as various current studies confirm [6,15].

Serious games have established a reputation for getting employees of companies involved in security activities in an enjoyable and sustainable way.

A. P. Fournaris et al. (Eds.): ESORICS 2019 Workshops, LNCS 11981, pp. 156–171, 2020.
https://doi.org/10.1007/978-3-030-42051-2_11

Moreover, serious games are designed for a primary purpose other than pure entertainment, e.g. education, awareness training or social change, but they preserve a playful character. Williams et al. [20] introduced the protection poker game to prioritise risks in software engineering projects. Shostack [18] from Microsoft presented his Elevation of Privileges card game to practice threat analysis with software engineers. Furthermore, games are used as part of security awareness campaigns [8] and particularly as a part of social engineering threat analysis [4].

Another game called *PERSUADED* specifically trains people to withstand social engineering attacks [1]. The game works as follows. Employees get confronted with a possible social engineering threat and have to select a defence mechanism. This correct defence mechanism is a pattern of behaviour ensuring a secure outcome. For example, an employee gets a phishing mail and is asked to open its attachment. Afterwards the player selects a countermeasure: "Do not open the email and inform the information security department immediately". The player gets immediate feedback whether the chosen defence is correct. In this paper, we describe how we built on the concept of PERSUADED and developed a new family of games called *PROTECT*.

Our contribution in this paper is the serious game PROTECT, which entails the following novelties:

- The game contains new scenarios for automated shipping and electronic cancer register domains.
- The game can be configured to serve various game settings to allow a progression between difficulty levels and various other challenges to keep the players playing.
- A discussion with five security practitioners to assess the potential of the game for security trainings.

Our paper is organised as follows. Section 2 presents background and related work, while Sect. 3 contains the design methodology applied for creating our game. Section 4 describes the serious game PROTECT in detail. Section 5 documents the feedback for the game from practitioners and Sect. 6 concludes.

2 Background and Related Work

As security is usually a secondary task, computer security training has often been perceived to be an uninteresting enforcement to users and managers. The approach of developing serious games has therefore been adopted to provide knowledge and training in that field.

CyberCIEGE is a role playing video game, where a player acts as an information security decision maker in an enterprise. Players' main responsibilities are to minimize the risk to the enterprise while allowing users to accomplish their goals. Similar to Persuaded, the game offers a simulation of the reality particularly portraying the need to maintain the balance between productivity and security. As decision makers, players get to make choices concerning users

(i.e. How extensive will background checks be?), computers (i.e. How will computers be networked?) and physical security (i.e. Who is allowed to enter a zone?) while monitoring the consequences of their choices. When compared to Persuaded, we recognized CyberCIEGE offered several advantages common to those offered by Persuaded. For instance, players are in a defensive mode and they get to make decisions and experience their consequences. CyberCIEGE even incorporates assets and resources in the game, which is a missing element in Protect. On the other hand, the game requires longer time to learn and to play [10].

PlayingSafe is a serious game in the domain of social engineering. It consists of multiple choice questions which are wrapped in typical mechanics of a board game. Since questions provided are exclusive to social engineering, the game is very similar to ours. The main difference lies however in the focus in the topic of social engineering. PlayingSafe asks questions in the fields of Phishing, advanced fee fraud, spam and others, being a category that covers less common attacks. Our game on the other hand covers a broader field without offering depth in each topic. Additionally, our game incorporates strategy favouring the entertainment element, in order to enhance the game experience the game provides [13].

SEAG is a serious game designed to raise awareness of social engineering. The game utilizes levels that tackle different cognitive aspects and hence provide an effective learning experience. The first level consists of quiz-like questions to build a knowledge base for the players. The second level is a match game where players have to match social engineering terms with respective pictures. Finally, the players are presented real life scenarios to analyse pertaining to threat. This simulation of real life application of the learnt lesson should test players ability to detect attacks- an approach very similar to inoculation [14]. Due to the construction with the different levels, the game seems to be more suitable for a one-time approach. In contrast, our game is based on one basic principle, but the configuration allows to raise the game's difficulty.

HATCH is a serious game for teaching employees about social engineering attacks [4]. The employees are guided by the game to elicit social engineering threats for their context. An extension of the game provides various scenarios e.g. for energy providers and personas to allow players to understand attacks of other contexts [5]. HATCH is a physical table top game that requires at least three players and a game master. Our game does not need a game master, and thus can be played by individual players at any time alone.

3 Methodology

PROTECT is based on the game concept of PERSUADED [1]. In this paper, Aladawy et al. discuss design goals and game concepts for a serious card game for the sensitization of people against social engineering attacks. To evaluate PERSUADED, a prototype implementation of the game has been developed.

It realizes the following improvements:

In this section, we describe the concepts for building PROTECT.

3.1 New Implementation with Enhanced Configuration

PROTECT is a complete new implementation of the design goals and game concepts of PERSUADED. While taking the findings from the case study into account, the focus was on the configuration of the game. By offering a lot of configuration options, i.e. for the game play, PROTECT can be seen as a family of games with PERSUADED just being a specific member of the game family. The aim is to allow an easy adaption to specific scenarios as well as to the player's skills. This can be particularly important if an employee changes the department and faces new threats in his/her new department.

By making the configuration options accessible via an application programming interface (API), PROTECT can not only serve as a stand-alone application but also be easily embedded into a training platform. In this case a training platform could control the difficulty of the game by changing the game configuration depending on the player's achievement in previous games. It would also be possible that the external training platform considers various other inputs such as the player's reaction to phishing mails, the results from other games or trainings.

In particular, we implemented an additional algorithm for the appearance of attacks in the game to make it easier for beginners to get started in the game. We introduced new cards that can defend any attack (jokers). We provided new algorithms for handling attacks which are not defended correctly and a special treatment for attacks that have not been defended correctly in previous games. The corresponding configuration parameters can be changed independently, allowing a number of (slightly) different games. The different configuration options are explained in detail in Sect. 4.2.

3.2 Game Concept

As for PERSUADED, the scientific foundation of this game are findings from Schaab et al. [16,17]. The authors analysed social psychology methods of training against persuasion and mapped them to trainings in IT security. One identified gap was the lack of using *inoculation*, the repeated confrontation of people with a challenging situation in order to trigger an appropriate response. In particular, inoculation is incorporated into the game mechanics to trigger resistance to social engineering attacks through exposing people to realistic attack scenarios. In order to provide the validity of the attack scenarios, we took all of them from scientific publications [2,7,11,12,15,19]. The game enables employees to learn about social engineering, while practising simultaneously. This immediate application of learned knowledge has proven to have lasting effects [9]. The enhanced configuration allows to adapt the game better to the player's needs. This is not only important to keep players motivated but also to adapt the game in a way that fits to the concept of inoculation. In versions for beginners the player's focus is mainly on matching different threats with the correct defences. In the more challenging versions for advanced players in order to be successful, the player is forced to think ahead. As a consequence, matching the different threats with the correct defences is still necessary but happens more unconsciously.

4 PROTECT

PROTECT is a serious card game that implements a training for the subject of social engineering. Its primary goal is the inoculation of people against social engineering attacks. This inoculation shall be achieved by confronting people repeatedly with social engineering scenarios in order to trigger an appropriate response.

PROTECT is implemented as an online game.

This chapter is divided into the following subsections:

- Section 4.1 describes game concepts and game mechanisms of PROTECT.
- Section 4.2 considers the configuration of PROTECT. In that respect, the configuration options regarding to (a) card decks, (b) instantiations of PROTECT and (c) properties for a game of PROTECT are discussed.
- Section 4.3 considers the implementation of PROTECT. It comprises the Graphical User Interface of PROTECT and its future provision as a web service.

4.1 Game Concepts and Game Mechanisms

This section considers the game concepts and mechanisms of PROTECT.

It is designed to achieve the following goals:

1. increasing awareness for social engineering,
2. training resistance to persuasion and
3. addressing the general population.

Regarding its main game concepts, PROTECT is designed as a single player card game that realizes a patience and solitaire game approach. As usual with this type of card games, the cards can be contained in the card deck or on the player's hand. In every turn of the game, a player can either draw a card from the deck or play a card from his/her hand. The implementation of these easy rules by PROTECT keep the complexity of the game low. This leads to a quite low initial barrier for playing the game and a focus on the actual content of teaching. Because the deck of cards is always shuffled before a game starts, each game is different from the previous game(s) (cf. [1], chap. 3, p. 5). This fact shall motivate players to play the game repetitively. The solitary approach enables players to play the game at any time, independently from other persons.

During a game of PROTECT, a player is confronted with different social engineering attacks. The task of the player is to select an appropriate defense mechanism for an attack. In this context, a defense mechanism represents a pattern of behaviour that prevents a successful conduct of a social engineering attack (cf. [1], chap. 1, p. 2). For the implementation of this game concept, PROTECT provides the following types of cards:

1. *Attack cards* represent scenarios for social engineering attacks in textual form.

2. *Defense cards* describe a pattern of behaviour for preventing the success of a certain attack. For each Attack card exists one corresponding Defense card. The contents of Defense cards are also represented in textual form.
3. *See The Future cards* allow the player to take a look on the three upper cards on the top of the card deck.
4. *Skip turn cards* allow the player to skip the top card of the deck and put this card to the bottom of the deck. It is only allowed to play a Skip turn card at the beginning of a turn when the top card of the deck is still hidden (cf. [1], Chap. 1, p. 4).
5. *Joker cards* are wildcards that can be selected by the player as a defence mechanism for every Attack card.

At the beginning of a game all cards are contained in the shuffled card deck. The game starts when the player draws the first card from the deck.

In the following, the game mechanisms of PROTECT are described.

At the beginning of a turn, a player can perform ONE of the following actions:

1. Draw a card from the top of the card deck.
2. Playing a See the future card or Skip turn card if such a card is on the player's hand.

Any drawn card that is NOT an Attack card, is put to the hand of the player. After that, the turn is over.

When an Attack card has been drawn, the player has to select the appropriate Defense card. If he/she

1. selects the correct Defense card, the score is increased.
2. selects an incorrect Defense card, the score is decreased and the player loses a life.
3. has no Defense card on the hand, a life is lost.

A player can also play a Joker card to defend every Attack card. In this case, the score is also increased. By playing Joker cards, players can achieve a good score, even if they do not know the appropriate defenses for some attacks. This shall keep up the motivation of the players high, to play the game repeatedly.

When the card deck is empty, the game is won. The game is lost if

1. the game time is up before finishing the deck or
2. a player has lost all his/her lives.

The following description considers the special function of See the future and Skip turn cards. As previously mentioned, it may be the case that a player has no appropriate Defense card or no Defense card at all on the hand to defend a drawn Attack card. If the player's hand does also not include a Joker card, he/she has no direct chance to prevent the loss of a life. This fact shall encourage the player to use See the future and Skip turn cards in the following way.

The player can play a See the future card to peek the upper three cards on top of the card deck. If these cards include any Attack cards, he/she can check if the appropriate Defense cards are

– on his/her hand or
– contained in the future cards itself at the right position.

If the future cards should contain any Attack cards for which no corresponding Defense card is available, the player can remember the order of these Attack cards and play a Skip turn card to skip such an Attack card when it is on the top of the deck. In this way, the loss of a life can be prevented. The provision of this game strategy increases the learning effect because the player studies the content of any Attack cards included in the future cards more carefully. This also applies for the content of the current Defense cards on his/her hand. Furthermore, he/she matches Attack cards partly against defense mechanisms that are not represented by Defense cards on the player's hand.

The provision of the strategy, mentioned before, requires an increased understanding of the game from the player. Additionally, it has a random factor because of the random order of the cards in the deck.

The study of [1] has shown that a considerable amount of players rated the above mentioned concept for the appearance of Attack cards on the top of the deck, as negative. Thus, PROTECT provides additionally a further concept for the appearance of Attack cards on the top of the deck. The implementation of this concept ensures that only such Attack cards can appear on the top of the deck for which an appropriate Defense card is currently on the player's hand. In this scenario the player can use the See the future cards and Skip turn cards to skip Attack cards for which he/she is not able to identify the appropriate Defense card on the hand. Because the additional concept for the appearance of Attack cards make the playing of PROTECT easier it shall be used for players on the beginner level.

PROTECT also provides two different concepts for the handling of Attack cards that have been solved incorrectly. In that regard, such an Attack card is

1. removed from the game or
2. is put back to the bottom of the card deck.

The second alternative represents the more easier variant because the player gets more chances to solve an attack correctly. Compared to the first variant, the player could still reach a good score, even with some incorrect solutions of attacks.

Example Scenario. We have extended the game PROTECT with various real scenarios from the EU project Threat Arrest[1]. One of these scenarios concerns automatic shipping. Digitalisation has increased the use of industrial control systems in the shipping domain. The increased use of computers and their interface exposes the systems that control vital systems and steer the ship itself to the risk of cyberattacks. The captain and crew are on their ship, while a back office provides IT-support. We elicited possible attacks that could be mitigated with awareness training such as the following. The crew is in contact with the back

[1] Threat Arrest homepage: https://www.threat-arrest.eu.

office on land in some intervals. If there is a problem with the onboard computer system the back office provides advice for maintenance to the crew. A social engineer pretends to be a back office employee and asks them to provide their credentials for maintenance. Another possible scenario would be that the crew is in ports all over the world. Maintenance is done on ports during stays outside of the home harbour. A social engineer pretends to be a maintenance worker and distributes usb sticks on the harbour with the hope that one of the crews picks one up and connects it to the computer system of the ship. We elicited totally over 20 plausible attacks for the game PROTECT.

4.2 Configuration Options

In this Section the options for the configuration of PROTECT are discussed. This discussion considers the following configuration aspects:

1. Configurations of card decks
2. Configurations during an instantiation of PROTECT
3. Internal configuration parameters of PROTECT

Configuration of Card Decks. Within PROTECT, the content of the cards of a deck are defined in a JSON format. Each card is defined by a single JSON file. The graphical representation of a drawn card in the GUI is generated on the fly during a game, based on the content of the corresponding JSON file. The definition of cards based on JSON files enables easy and fast

- creations of new card decks and
- modifications of existing card decks

to cover more specific social engineering scenarios.

Each card deck in PROTECT is identified by an unique identifier. These identifiers are used to configure which card deck shall be played within an instantiation of PROTECT (see Sect. 4.2).

Standard Card Decks. The standard card deck of PROTECT contains pairs of Attack and Defense cards for typical social engineering scenarios. It includes the following types off attacks (cf. [1], chap. 1, p. 4):

- baiting,
- phishing,
- tailgating
- mail attachment,
- physical impersonation,
- virtual impersonation,
- voice of Authority and
- popup window.

Additionally, the standard card deck contains action cards in form of Joker, See the future and Skip turn cards. The number of action cards of each type in the card deck can be configured when PROTECT is instantiated (see Sect. 4.2).

Adapted Card Decks. The game PROTECT can be also used to verify that a company's security policy is understood and followed by its employees. This works by describing the possible attacks against a company that the rules of the policies try to prevent. For example, the policy might contain a rule to shredder all confidential documents. We provide a card in which a person takes the shredder for maintenance and tells the staff that in the absence of the shredder they should throw the documents in the regular trash bin and that is not necessary to use the shredder on the next floor. The right behaviour would be to object and use the other shredder and inform the security staff of this incident.

Instantiation Parameters. PROTECT provides the hand over of information that is necessary for a game, during its instantiation. This information is represented by so-called *instantiation parameters* that are listed in Table 1.

The instantiation parameters *player ID* and *player name* provide information about the player of the game. The time that a game can take the longest is represented by the parameter *game time*. Because of their logical connection the instantiation parameters *card deck ID* and *difficulty level* shall be considered in more detail. The *card deck ID* and *difficulty level* enable the definition which card deck shall be played with which level of difficulty. Within PROTECT, a value for a difficulty level is mapped to a certain configuration of PROTECT regarding the selected card deck. This means, that a level of difficulty results from the particular values of the configuration parameters. These configuration parameters are specified in Table 2.

The parameter *special practice* defines if Attack cards that have been solved incorrectly in previous rounds of the game and there corresponding Defense cards shall be included multiple times in the card deck. If this is the case, the number of occurrences for such pairs of cards is defined by the appropriate configuration parameter (see Table 2).

Table 1. Instantiation parameters of PROTECT

Parameter	Description
player ID	Unique identifier of the player
player name	Name of the player
game time	Game time in minutes
card deck ID	Unique identifier of the card deck that shall be played
difficulty level	Level of difficulty with which the game shall be played. The value of the difficulty level corresponds to a certain internal configuration of PROTECT
special practice	Defines if Attack cards that have been solved incorrectly in previous games of PROTECT and the appropriate Defense cards shall occur multiple times in the card deck

Internal Configuration Parameters. *Internal configuration parameters* enable a configuration of properties for a game of PROTECT. The different internal configuration parameters are described in Table 2. A set of internal configuration parameters with the appropriate value is contained in a *configuration*. Configurations specify certain levels of difficulty for a game of PROTECT by the values of their parameters. For example, the level of difficulty decreases

- the more Joker cards a card deck includes,
- the more lives a player has,
- when only such Attack cards can be drawn for which the corresponding Defense card is on the player's hand,
- when incorrectly solved Attack cards are put back into the card deck and
- when the score can not have a value less than zero.

A configuration is associated to a certain difficulty level for a play of PROTECT with a particular card deck. The information according to the card deck and difficulty level are passed during the instantiation of PROTECT (see Table 1).

Table 2. Internal configuration parameters of PROTECT

Parameter	Description
number of lives	Defines the numbers of lives that a player has
number Joker cards	Defines the number of Joker cards in the card deck
number See the future cards	Specifies the number of See the future cards in the card deck
number Skip turn cards	Defines the number of Skip turn cards in the card deck
score increase	Defines the number of points added to the score when the CORRECT Defense card or a Joker card has been selected for an Attack card
score decrease	Defines the number of points removed from the score when an INCORRECT Defense card has been selected for an Attack card
range of score	Specifies if the score can be less than zero or if the lowest score is zero
appearance of Attack cards	Defines if (a) ANY Attack card can appear on the top of the deck, even if the corresponding Defense card is not on the hand of the player. (b) ONLY those Attack cards can appear on the top of the card deck for which the corresponding Defense card is on the player's hand
handling of incorrectly solved Attack cards	Specifies if an Attack card that has been solved incorrectly is (a) put back to the bottom of the card deck or (b) removed from the game

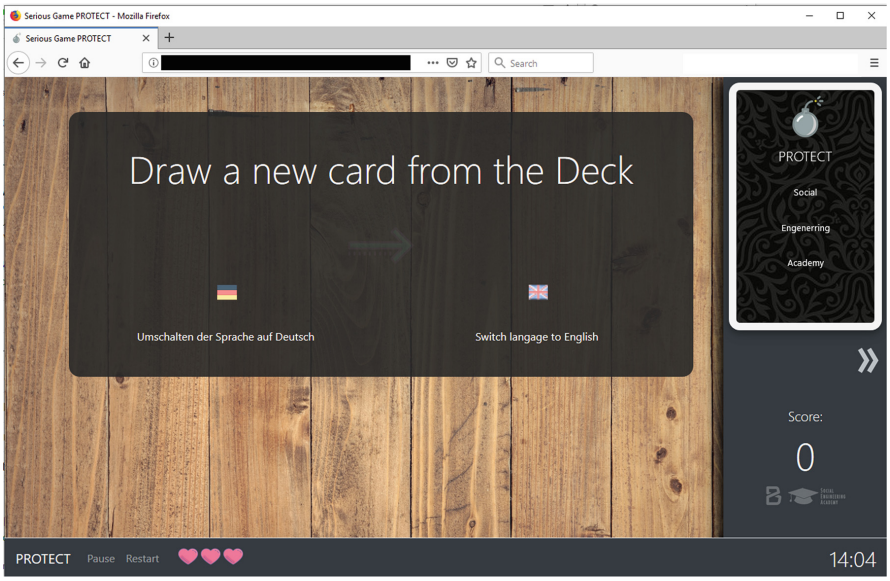


Fig. 1. GUI of PROTECT at the beginning of a game

4.3 Implementation

This Section discusses the implementation of game concepts and mechanisms that are described in Sect. 4.1 by PROTECT. The discussion considers the Graphical User Interface of PROTECT and a concept for its future provision as a web service.

Graphical User Interface. The Graphical User Interface (GUI) of PROTECT is executed in a web browser. It is implemented in JavaScript by using the JavaScript library *jQuery* and the framework *Bootstrap*. The GUI is especially designed to be displayed on mobile devices. Nonetheless, it can be displayed on PC monitors and laptop screens without any problems.

The Fig. 1 shows an execution of the PROTECT GUI in a web browser at the beginning of a game. The dialog for changing the language of the game is displayed. The card deck, including the Attack, Defense and action cards (e.g See the future cards), is positioned in the top right corner. It is shuffled automatically before each game. A player can draw a card by double-clicking on the card deck. The game score and the remaining game time are represented in the bottom right corner. It is also possible to pause a game with help of the *Pause*-button in the bottom left corner of the GUI. A game can also be cancelled and restarted. The corresponding *Restart*-button is positioned next to the *Pause*-button. The remaining lives of a player during a game are displayed by the pink heart symbols.

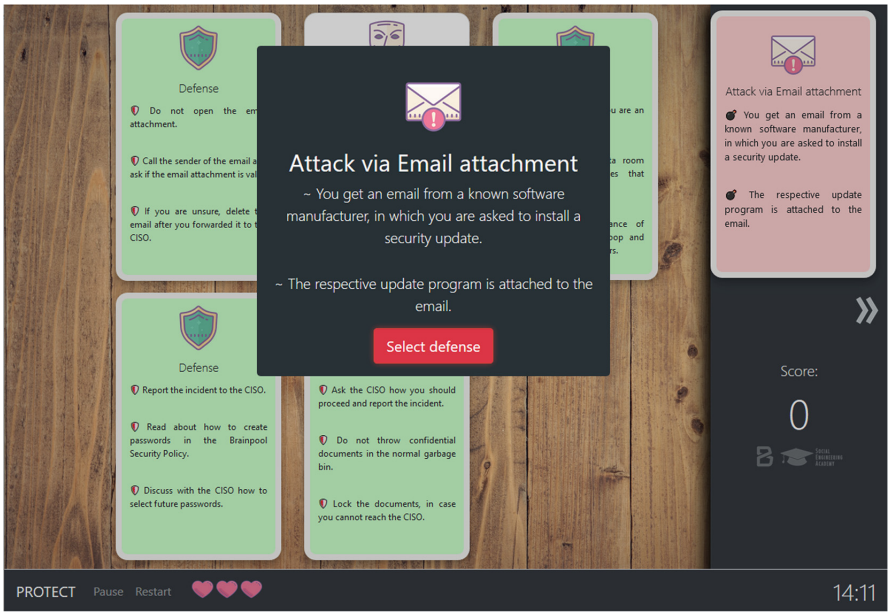


Fig. 2. Dialog after an Attack card has been drawn

The GUI supports the player in the game flow with appropriate dialogs. For example, the dialog in Fig. 2 is shown after the player has drawn an Attack card. It requests the player to select a defense card after clicking on the *Select defense*-button. The Fig. 3b displays the dialog after the selection of the correct Defense card. The game continues, after the player has pressed the *Continue*-button.

For example, the dialog in Fig. 3a is shown after the player has drawn an Attack card. It requests the player to select a defense card after clicking on the *Select defense*-button. The Fig. 3c displays the dialog after the selection of an incorrect Defense card. When the player clicks on the *Show the right answer*-button, the subsequent dialog represents the correct defense mechanism for the drawn Attack card (see Fig. 3d). The game continues, after the player has pressed the *Continue*-button.

Provision of PROTECT as a Web Service. The content of this section describes a concept for the provision of PROTECT as a web service. This type of provision has the following advantages:

1. Companies that want to use PROTECT for training their employees do not need to set up an own infrastructure for the deployment of PROTECT.
2. PROTECT can be integrated easily into other training platforms. This is achieved by the use of standardized application protocols that enable a loose coupling between different systems. Such an approach will be realized within

(a) Attack card has been drawn (b) Selection of the correct Defense card

(c) Selection of an incorrect defense card (d) Correct defense mechanism

Fig. 3. Different dialogs within the game

the research project *Threat-Arrest*[2], where PROTECT will be integrated into the *Threat-Arrest training platform*.

PROTECT shall be provided as a cloud computing service in form of *Software as a Service (SaaS)*. For the deployment of PROTECT, an appropriate cloud infrastructure, deployment environment and database shall be used in form of cloud services. The usage of these services shall be supplied by a third party cloud service provider.

The architecture of the PROTECT web service is represented in Fig. 4 in an abstract way. It shows that the PROTECT web service will use

- a deployment environment for the deployment of PROTECT and
- a data base service for storing data that is related to played games of PROTECT.

The selection of a certain deployment environment (e.g. virtual server, container service) is currently in the state of development.

[2] https://www.threat-arrest.eu/.

The external functionality of the PROTECT web service is provided via a REST API (see Fig. 4). A client can use a certain function by sending the appropriate HTTP request to the PROTECT web service. The web service sends the result of the function back to the client via an HTTP response. In the following, the basic functionality of the PROTECT REST API will be considered:

1. Instantiation of PROTECT with the specified instantiation parameters (see Table 1). The PROTECT web service returns the created PROTECT instance to the client.
2. Query of results regarding to games of PROTECT. The set of the returned results can be defined by filter parameters that are contained in the content data of the appropriate HTTP requests.

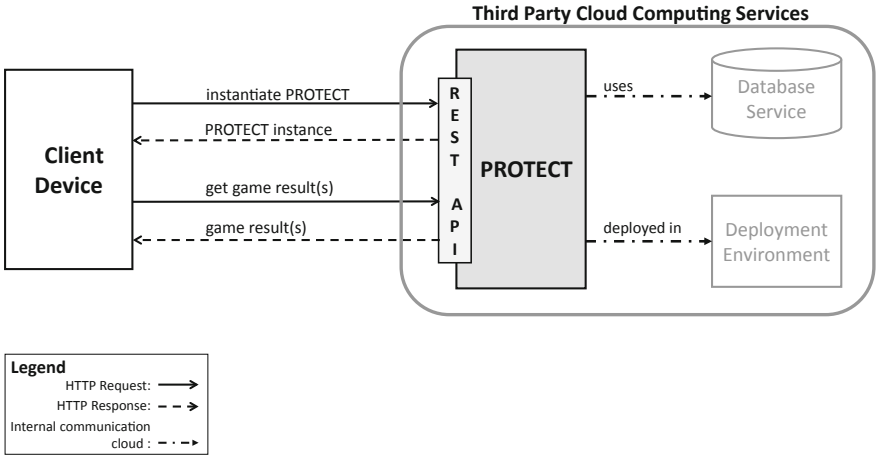


Fig. 4. Abstract architecture of the PROTECT web service

5 Discussion

We showed our game to 5 practitioners from different domains (from the information technology, cybersecurity, smart homes, and automotive) and gathered the following feedback. The general perceptions during game play provided feedback that conforms with our design goals. One participant stated, that for them the game simulates the reality. The player further explained, that in real life, it is rather difficult to expect social engineering attacks and always be ready for them, which they found was mapped through the random factor. Furthermore, the player mentioned that usually even the most cautious people might fall victim for social engineering when not constantly reminded of this threat. The player stated that the game does a good of job of doing that. Moreover, being

able to defend themselves against social engineering in the game gave confidence the same could be achieved in real life. Another participant provided feedback on the challenge level in the game saying that "one has to think". For this player, the game was also "easy to understand", reflecting the modesty of the trade-off between those two conflicting elements. Finally, the player emphasised the importance of the game being single player for the replay value, saying that he can "play the game another 3 times just right now".

6 Conclusion

We designed, implemented and evaluated a serious game family for training social engineering defence mechanisms, called PROTECT. Since the basic concept of the game has already been evaluated for PERSUADED [1], we focused on the evaluation of the enhanced configuration.

Several goals were specified and refined to achieve the serious purpose of the game:

- Easier start into the game and increased replay probability.
- The game, i.e. game play, can be adapted to the player's skills and previous game results.
- The game attack scenarios can easily be adapted to the player's skills and environment.
- An integration into external training platforms is allowed, i.e. the platform can decide about the difficulty of the next games.

Our qualitative evaluation showed that with the enhanced configuration options, we could achieve our purpose. In future work, we aim to do a quantitative evaluation with a larger number of players.

Acknowledgements. This work has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 786890 (THREAT-ARREST).

References

1. Aladawy, D., Beckers, K., Pape, S.: PERSUADED: fighting social engineering attacks with a serious game. In: Furnell, S., Mouratidis, H., Pernul, G. (eds.) TrustBus 2018. LNCS, vol. 11033, pp. 103–118. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-98385-1_8. ISBN 978-3-319-98384-4
2. Bakhshi, T., Papadaki, M., Furnell, S.: A practical assessment of social engineering vulnerabilities. In: HAISA, pp. 12–23 (2008)
3. BBC: How to hack people (2002). news.bbc.co.uk/2/hi/technology/2320121.stm
4. Beckers, K., Pape, S.: A serious game for eliciting social engineering security requirements. In: Proceedings of the 24th IEEE International Conference on Requirements Engineering (RE 2016). IEEE Computer Society (2016). https://doi.org/10.1109/RE.2016.39

5. Beckers, K., Pape, S., Fries, V.: HATCH: hack and trick capricious humans - a serious game on social engineering. In: Proceedings of the 2016 British HCI Conference, 11–15 July 2016, Bournemouth, United Kingdom (2016). http://ewic.bcs.org/content/ConWebDoc/56973
6. Dimensional Research: The Risk of Social Engineering on Information Security: A Survey of IT Profesionals (2011). http://docplayer.net/11092603-The-risk-of-social-engineering-on-information-security.html
7. Ferreira, A., Coventry, L., Lenzini, G.: Principles of persuasion in social engineering and their use in phishing. In: Tryfonas, T., Askoxylakis, I. (eds.) HAS 2015. LNCS, vol. 9190, pp. 36–47. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-20376-8_4
8. Gondree, M., Peterson, Z.N.J., Denning, T.: Security through play. IEEE Secur. Priv. **11**(3), 64–67 (2013)
9. Greitzer, F.L., Kuchar, O.A., Huston, K.: Cognitive science implications for enhancing training effectiveness in a serious gaming context. J. Educ. Resour. Comput. **7**(3), 2 (2007)
10. Irvine, C.E., Thompson, M.F., Allen, K.: CyberCIEGE: gaming for information assurance. IEEE Secur. Priv. **3**(3), 61–64 (2005)
11. Manske, K.: An introduction to social engineering. Inf. Syst. Secur. **9**(5), 1–7 (2000)
12. Mitnick, K.D., Simon, W.L.: The Art of Deception: Controlling the Human Element of Security. Wiley, Hoboken (2011)
13. Newbould, M., Furnell, S.: Playing safe: a prototype game for raising awareness of social engineering. In: Australian Information Security Management Conference, p. 4 (2009)
14. Olanrewaju, A.S.T., Zakaria, N.H.: Social engineering awareness game (SEAG): an empirical evaluation of using game towards improving information security awareness. In: Proceedings of the 5th International Conference on Computing and Informatics (ICOCI 2015) (2015)
15. SANS: Social Engineering Threats (2003). http://www.sans.org/reading-room/whitepapers/engineering/threat-social-engineering-defense-1232
16. Schaab, P., Beckers, K., Pape, S.: A systematic gap analysis of social engineering defence mechanisms considering social psychology. In: Proceedings of the 10th International Symposium on Human Aspects of Information Security & Assurance (HAISA 2016), 19–21 July 2016, Frankfurt, Germany (2016). http://www.cscan.org/openaccess/?paperid=301
17. Schaab, P., Beckers, K., Pape, S.: Social engineering defence mechanisms and counteracting training strategies. Inf. Comput. Secur. **25**(2), 206–222 (2017). https://doi.org/10.1108/ICS-04-2017-0022
18. Shostack, A.: Threat Modeling: Designing for Security, 1st edn. Wiley, Hoboken (2014)
19. Stajano, F., Wilson, P.: Understanding scam victims: seven principles for systems security. Commun. ACM **54**(3), 70–75 (2011). https://doi.org/10.1145/1897852.1897872. http://doi.acm.org/10.1145/1897852.1897872
20. Williams, L., Meneely, A., Shipley, G.: Protection poker: the new software security "game". IEEE Secur. Priv. **8**(3), 14–20 (2010)

A.8 Systematic Scenario Creation for Serious Security-Awareness Games

Vera Hazilov and Sebastian Pape. Systematic scenario creation for serious security-awareness games. In Ioana Boureanu, Constantin Câtâlin Drâgan, Mark Manulis, Thanassis Giannetsos, Christoforos Dadoyan, Panagiotis Gouvas, Roger A. Hallman, Shujun Li, Victor Chang, Frank Pallas, Jörg Pohle, and Angela Sasse, editors, *Computer Security - ESORICS 2020 International Workshops, DETIPS, DeSECSys, MPS, and SPOSE, Guildford, UK, September 17-18, 2020, Revised Selected Papers*, volume 12580 of *LNCS*, pages 294–311, Cham, 09 2020. Springer International Publishing. doi: 10.1007/978-3-030-66504-3_18. URL https://link.springer.com/chapter/10.1007/978-3-030-66504-3_18

Systematic Scenario Creation for Serious Security-Awareness Games

Vera Hazilov[1] and Sebastian Pape[2,3](✉)

[1] Intero Operations and Services GmbH (INOS), Munich, Germany
[2] Chair of Mobile Business and Multilateral Security, Goethe University Frankfurt, Frankfurt, Germany
sebastian.pape@m-chair.de
[3] Social Engineering Academy GmbH, Frankfurt, Germany

Abstract. While social engineering is still a recent threat, many organisations only address it by using traditional trainings, penetration tests, standardized security awareness campaigns or serious games. Existing research has shown that methods for raising employees' awareness are more effective if adjusted to their target audience. For that purpose, we propose the creation of specific scenarios for serious games by considering specifics of the respective organisation. Based on the work of Faily and Flechais [11], who created personas utilizing grounded theory, we demonstrate how to develop a specific scenario for HATCH [4], a serious game on social engineering. Our method for adapting a scenario of a serious game on social engineering resulted in a realistic scenario and thus was effective. Since the method is also very time-consuming, we propose future work to investigate if the effort can be reduced.

Keywords: Serious game · Security awareness · Personas · Scenario creation

1 Introduction

Social engineering is older than the electronic age itself and is still a part of our life. The European Network and Information Security Agency, ENISA, defines social engineering as a technique that exploits human weaknesses and aims to manipulate people into breaking normal security procedures [21]. In most cases, maliciously motivated attackers aim to gain access to their victims' commercial, financial, sensitive or private information in order to use it against them or cause harm otherwise [2]. Social engineering's key elements are deception, exploitation and use of psychological tricks. Social engineering attacks represent a threat to individuals and organisations and often lead to some kind of financial losses.

However, most organisations have difficulties addressing this issue adequately. According to Kevin Mitnick – a former hacker who now works as an IT security consultant, most companies rather purchase heavily standardized security products, such as firewalls or intrusion detection systems, than considering potential threats of social engineering attacks [19]. Mitnick criticizes this approach

I. Boureanu et al. (Eds.): ESORICS 2020 Workshops, LNCS 12580, pp. 294–311, 2020.
https://doi.org/10.1007/978-3-030-66504-3_18

and argues that technology-based products simply create an illusion of security however, leave organisations disarmed towards attacks that are directed towards their employees. Peltier [22] supports this argument and states that technology-based countermeasures should be applied whenever possible. However, he also claims that no hardware or software is able to protect an organization fully against social engineering attacks. In addition to that, social engineering is highly interdisciplinary, however most defense strategies are advised by IT security experts who rather have a background in information systems than psychology [26,27].

Traditional trainings mainly focus on transfer of knowledge and often do not address employees' attitude towards security or raise their awareness sufficiently. While knowledge is a prerequisite to counter social engineering attacks, a successful defense also requires a sufficient security-aware culture among staff [1], which represents a challenge for many organisations. Mainly because security policies are often in a bad shape and rather inform employees about what not to do than providing any guidance about desired behaviour and outcomes. Penetration tests are attached to a lot of obligations and legal burdens that need to be resolved beforehand. They can demotivate employees, who as a consequence might give up on defending social engineering attacks at all, and usually can not be repeated regularly, because employees become aware of penetration testers [9]. Security awareness campaigns often fail because they evoke negative feelings such as anxiety, fear or stress and are therefore often ineffective. In addition to that, individuals generally dislike following advice or instructions because it is associated with losing control. Lastly, awareness campaigns often provide only information about risks, are often not engaging, interesting and entertaining enough and therefore fail to change individuals' behavior [3]. Serious games however, are more entertaining and engaging than traditional forms of learning and can influence individuals' behavior due to their use of pedagogy and game-based learning principles, such as motivation, cognitive apprenticeship and constructivism [10]. They have demonstrated a potential in industrial education and training disciplines [23,25] if respective organizations care for players' privacy and working atmosphere [16], do not use gaming data for appraisal or selection purposes and clearly communicate this to the employees [17]. Abawajy's observations [1], that trainings can be greatly enhanced through interactive content, support this statement and make serious games a strong candidate for overcoming issues of traditional training methods.

However, not only for security awareness campaigns, but also for serious games it is important to address the target audience as specific as possible. Therefore, in this paper, we aim to adjust a serious game to a specific target group by adapting it accordingly. For that purpose we chose the serious game HATCH [5] and developed a new scenario for one of its variants in order to be suitable for consulting companies. This approach tackles that problem, that although many serious games for IT security exist, it is still hard to find a accurately fitting serious game for a specific organisation or scenario.

2 Background and Related Work

This work is based on two concepts, personas and HATCH. Personas represent a popular technique that is often used in user-centered design in order to create services, products or software [24]. HATCH is a serious game on social engineering, for which we have developed a scenario as proof of concept. However, hardly any specific properties of the game were used, so it should be possible to generalise the results and develop scenarios for related games.

2.1 Personas

By definition, personas are imaginary however, realistic descriptions of stakeholders or future users of a service or product, who have names, jobs, feelings, goals, certain needs and requirements [11]. The concept was firstly introduced by Cooper [7] in 1999. Cooper argues that developers need to consider future users' needs, goals and wishes, instead of designing products for 'elastic users'. The latter term represents highly standardized descriptions of users, which are unrealistic and in many cases rather represent developers' own needs. According to Cooper, the use of elastic users therefore leads to products, which only partly satisfy real users' needs.

In 2011, Faily and Flechais [11] introduced a method for developing personas that is based on grounded theory. The latter is a "[...] systematic, yet flexible guideline for collecting and analyzing qualitative data" [6]. Faily and Flechais [11] collected necessary data through interviews, each of them lasting approximately an hour. All interviews have been transcribed and subjects to a grounded theory analysis using ATLAS.ti, a qualitative data analysis and research tool. The process of developing personas included three steps [11]: the first step includes reading all interview transcripts, identifying relevant text passages, assigning appropriate phrases (codes) to them and formulating them as propositions. The propositions are later summarized and as a result represent most significant concepts developed personas need to explore. As next, appropriate propositions are selected and stated as potential characteristic of a persona. The final step of this approach involves selecting relevant characteristics and writing a persona narrative. Faily and Flechais [11] used their approach successfully to derive accurate archetypes of their respective user communities (personas) from around 300 quotations and 90 thematic concepts.

2.2 HATCH

Hack and Trick Capricious Humans (HATCH) is a physical (tabletop) serious game on social engineering [4,5]. The game is available in two versions, a real life scenario and a generic version. Each version of the game pursues a slightly different objective: The real life scenario is aiming to derive social engineering security requirements of a company or one of its departments. Therefore, a real environment is modelled and players attack their colleagues in order to identify real attack vectors. The generic version of the game aims to raise players' awareness

for social engineering threats and educate them on detecting this kind of attacks. In order not to unnecessarily expose and blame colleagues during a training session, it is based on a virtual scenario with personas as attack victims [16]. The scenario consists of a layout of a medium-sized office and ten personas, which are fictional descriptions of employees. All of which are printed on cards and contain information such as this employee's name, role, familiarization with computers and attitude towards security and privacy [5].

In both versions two deck of cards are used (psychological principles and social engineering attacks). When playing the game, each player draws one psychological principle card and three social engineering attack cards and reads the respective descriptions. Psychological principle cards state and describe human behaviors or patterns that are often exploited by social engineers, as for example: 'Distraction - While you distract your victims by whatever retains their interests, you can do anything to them'. On the other hand, the social engineering cards name and define some of the most common social engineering attacks, for example dumpster diving, which is 'the act of analyzing documents and other things in a garbage bin of an organization to reveal sensitive information'. Each player has then the task to choose a victim[1] which fits to the psychological principle card and elaborate an attack by using one of the social engineering attack cards which matches the victim and psychological principle best.

Players take turns to reveal their cards and describe the social engineering attack they came up with. Other players discuss the proposed attack and award points for attack's feasibility and viability and rate if it is compliant with descriptions of this player's cards. The total score of each player is calculated by the end of the group rating and the player with the highest score wins the game. At the end of the game, all players briefly reflect on proposed social engineering attacks and derive potential security threats.

Beckers and Pape [4] showed that the real life scenario was helpful to increase the security awareness of employees [5] and in the elicitation of context-specific attacks by utilizing the domain knowledge of the players and their observations and knowledge about daily work and processes.

3 Methodology

The data that was used to develop a consulting services scenario for HATCH was collected through expert interviews, which have proven to be of good practical value [18]. The interviews were executed as semi-structured interviews based on the interview guide described in Sect. 3.1. Section 3.2 describes the interviewees and Sect. 3.3 the subsequent coding and qualitative analysis.

3.1 Interview Guide

Meuser and Nagel [18] emphasize the importance of using an interview guide. In particular for semi-structured interviews they serve two purposed. On the one

[1] Depending on the version either a colleague or a persona.

hand, they help the interviewer to not get lost in irrelevant topics and focus on the goal of the interview [12]. On the other hand, they help the interviewer to organize and structure the interviews and adapt them to knowledge gained in previous interviews [20].

The interview guide was constructed taking following aspects into consideration:

- the appropriate number of questions – although a large number of questions might provide deeper insights, too many questions can also extend the interview to an inefficient level. In alignment the suggestion from Gläser and Laudel [13] to limit the number of questions to approximately fifteen, the derived interview guide consists of seventeen questions.
- appropriate format of questions – asked questions can be noted as fully formulated sentences which provides stability or stated vaguely which increases interviewer's flexibility to react ad hoc [13]
- appropriate content of questions, which means that asked questions can be based on existing theories, publications or interviewer's own experience or knowledge [12].

The interview guide was tested within two one-hour interview sessions. At the end of each session, interviewed experts were asked to provide feedback regarding the guide's length, format and content. The initial interview guide was adopted during the process based on received feedback: an explanation of this work's main objective and approach was added to the introduction section. The interview guide's second section was extended by a definition of the term social engineering for the purposes of general introduction. All remaining sections stayed unchanged and aim to uncover this industry's specifics, assets, communication channels, their physical location as well as existing roles, skills and attitudes towards security and privacy. Table 1 gives a brief overview of the interview guide's structure.

3.2 Interview Implementation and Participants

All nine expert interviews were conducted in January and February 2017 and lasted between 35 min and 61 min (cf. Table 2). All interviews were conducted in German –the experts' native language in order not to obstruct experts' thinking ability and allow them to provide complex and comprehensive answers. Most interviews were conducted face-to-face, only interview seven and eight were recorded over the phone. None of the participants received any printed information, such as handouts or printouts, before or during the interview in order to avoid any distraction. However, before the interviews, participants were informed about the study's approach and goal and asked for consent as indicated in Table 1. Table 2 presents an overview of all participants, their role, professional experience, corresponding business unit and the interview's duration.

Due to difficulties of cold calling professional consultants and requesting their help for creating a serious game scenario, all participating interviewees were approached based on existing contacts. Furthermore, none of the approached

Table 1. Interview guide

#	Section	Content
1	Introduction	• Greeting and opening • Statement of classification • Declaration of consent • Introduction to the research's approach and main goal
2	Social engineering	• General understanding • Definition • Previous experience with SE attacks
3	Industry's specifics	• General understanding • Associations • Characteristics
4	Assets & location	• Company's assets and employees • Asset's location
5	Roles & tasks	• Specific roles • Responsibilities and tasks
6	Communication channels	• Company's communication channels • Management process • Access rights • Relevant content
7	Personas	• Skills • Knowledge • Attitude towards security and privacy

employees of a 'client company' were willing to participate, since they were afraid of revealing sensitive information which could potentially lead to a social engineering attack. However, we do not think that this was a major drawback, since the developed scenario aimed to focus on consulting companies.As a consequence, all interviewed experts have in common that they are employed by a large consulting/auditing firm, however differ in their roles, business units, gender, age and level of professional experience. The experts' selection was done in order to introduce a certain level of variety, however contain a strong focus at the same time: We expected that a more unified selection of participants would have resulted in a highly specific scenario, while a too diverse selection of experts may have yielded unfocused results.

3.3 Data Analysis

All interviews were audio recorded, transcribed literally[2], and all transcripts were imported into MAXQDA, a professional software for qualitative text analysis, and coded in chronological order. The applied process of coding consisted of two rounds, open and axial coding. While open coding is the process of reading

[2] Pauses and certain sounds were neglected such as 'huh' etc.

Table 2. Participants overview

#	Role	Experience	Business unit	Duration
1	Consultant	1–3 years	Management consulting	61 min
2	Consultant	1–3 years	Risk consulting	54 min
3	Consultant	6+ years	Technology consulting	55 min
4	Consultant	3–6 years	Technology consulting	35 min
5	Assistant	1–3 years	Management consulting	37 min
6	IT	1–3 years	Technology consulting	60 min
7	Consultant	3–6 years	Technology consulting	35 min
8	Consultant	6+ years	Technology consulting	59 min
9	Consultant	6+ years	Technology consulting	46 min

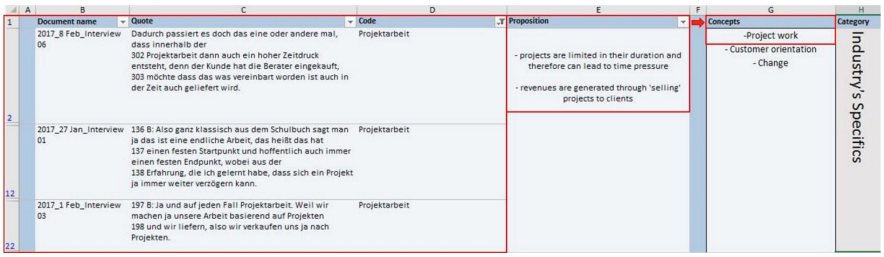


Fig. 1. Process of axial coding

textual data line-by-line, identifying certain phenomena within it and attaching adequate phrases (e. g. codes) to it, axial coding represents the process of examining previously assigned codes, identifying certain relationships among them and summarizing them into concepts and categories [8]. This work's coding process is illustrated in Fig. 1.

The illustration above shows a fraction of all text passages that have been assigned with the code 'project work' and later formulated into propositions 'projects are limited in their duration and therefore can lead to time pressure' and 'revenues are generated through selling projects to clients'. All relevant propositions were later summarized to the concept 'project work' and assigned to the category 'industry's specifics'. Following this approach, 110 pages of interview transcripts were assigned with 509 codes.

3.4 Development of the Scenario

Since we took HATCH for granted, as it already existed before, we do not describe its development, however focus on the creation of a new scenario. Figure 2 illustrates the steps of the scenario development.

In the previous sections, we have already described the interview, transcript and coding phases (stage 1 to 3). Following Faily and Flechais' method [11] for developing personas, we developed propositions from codes (stage 4), such as ' more consultants are hired for project than clients', 'with the exception of client's assistants, consultants are generally younger' and 'generally, the consulting team consists of 4 to 5 people'. These propositions were summarized, assigned to concepts and categorized (stage 5). For example, previous propositions were assigned to the concepts 'role' and 'age' and categorized as 'personas'. Altogether 21 concepts were sorted into five categories, which represent the main components of the consulting services scenario for HATCH. Those categories are: industry's specifics, assets, communication channels, location and personas. The first four of them represent a consulting firm's working environment, while the last embodies personas' characteristics.

As the last step, appropriate propositions were selected and stated as potential characteristics of a persona to write persona narratives and develop the scenario (stage 6). For this purpose, all personas-related concepts and propositions were reviewed again, in order to identify most valuable and meaningful insights, and later embodied into future personas. For example, the propositions from the concepts 'roles' and 'age' lead to the decision of having more consulting personas (4) than personas of the client company (3). Furthermore, with the exception of the client's assistant, all consulting personas are younger than personas of the client. In the same manner, propositions were used to develop professional consultants' working environment and surroundings.

3.5 Evaluation

Note that this work focuses on creating a new scenario in order to adapt an existing game called HATCH, which has already been evaluated [4,5]. Therefore, we did not evaluate the game, its rules and elements itself, however rather focused on evaluating the consulting services scenario.

The developed consulting services scenario for HATCH was evaluated by five players and within two sessions: the first session was conducted on 30th of March 2017 and lasted roughly one hundred minutes, while the second session took place on March 31st, 2017 and continued approximately two hours. One moderator was present at both sessions and all players had an IT background, were employed by an auditing/consulting firm. None of the players was involved in the previous interview sessions.

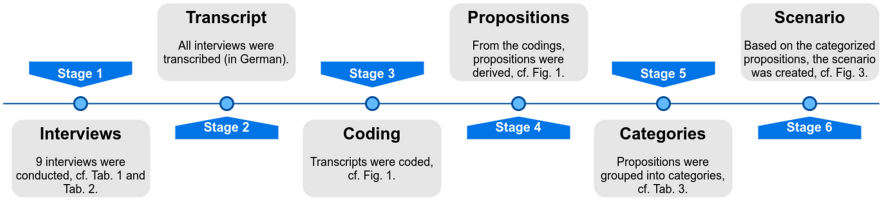


Fig. 2. Overview of scenario creation process

Table 3. Derived scenario related categories and concepts

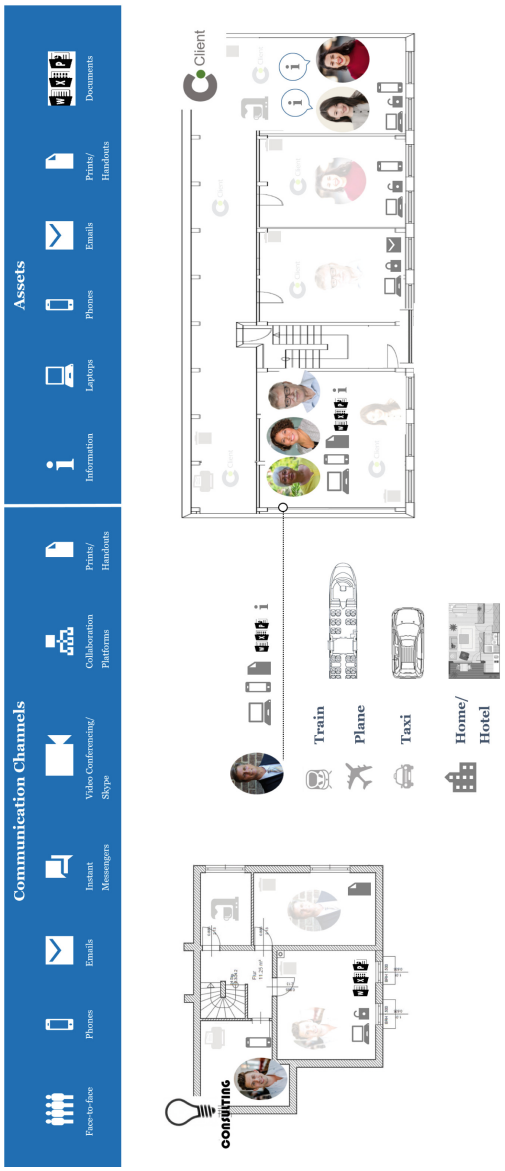
#	Category	Concept
1	Industry's specifics	• Project work • Customer orientation • Change
2	Assets	• Information (sensitive, project-related, private) • Laptops • Phones • Emails • Prints, handouts • Documents (office)
3	Communication channels	• Face-to-face • Phone calls • Emails • Video conferences • Collaboration platforms • Prints, Handouts
4	Location	• Client's office • Remote locations
5	Personas	• Age • Roles and tasks • Skills and knowledge • Attitude towards security and privacy

4 Results

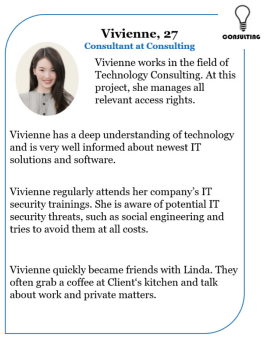
With the process described in the previous section, we derived five relevant categories with altogether 21 concepts as shown in Table 3. The industry's specifics, consultants' assets, communication channels and location are incorporated within the scenario, which represents working environment and surroundings of professional consultants. These companies' assets and communication channels are pictured at the top of the scenario, since their location might vary a lot between companies and we aimed to avoid a too strict mapping to an individual or a certain location (cf. Fig. 3a). The results from the personas category were used to create different persona cards as shown in Fig. 3b to Fig. 3d.

4.1 Scenario

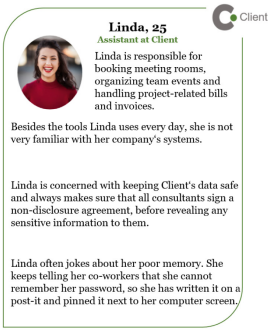
Besides the layout of both companies, called Consulting and Client, the scenario represents this industry's characteristics and includes several personas, which are described in the next section. As illustrated in Fig. 3a, consulting firms use a number of communication channels, such as face-to-face interaction, phones, emails, instant messengers, video conferencing tools or Skype, collaboration platforms,



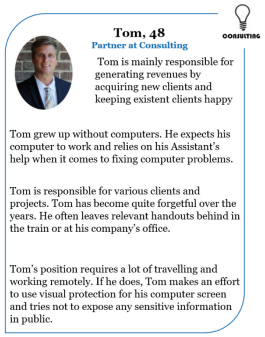
(a) Scenario



(b) Vivienne, Consultant at Consulting



(c) Linda, Assistant at Client



(d) Tom, Partner at Consulting

Fig. 3. Scenario and personas "consulting company"

prints, handouts and posses assets that are mostly focused around information: laptops, phones, emails, prints or handouts and Word, Excel or PowerPoint documents.

One characteristic of firms within this industry is project-based work. Consulting companies generate revenues by selling their services in form of projects, which are mostly executed at their customers' office. Therefore, consultants are required to travel a lot and work from various locations e.g. their own or client's office, public transportation, hotel rooms or from home. Therefore, the presented scenario pictures personas in various locations, including layouts of two offices, which contain several elements and details. Consulting's office is placed on the left, it has a kitchen and several rooms, while Client's office is pictured on the right.

4.2 Personas

The scenario also contains seven personas, which are fictional descriptions of workers that are employed by the consulting company or the organisation that hired them. All personas include information such as an employee's name, age, occupied role, tasks, attitude towards security/privacy and personality traits. Players will get cards with the description of the personas as shown in Fig. 3b to Fig. 3d. We also provide a more schematic presentation in Table 4. Since both presentations can not describe the interactions, the remainder of this section describes developed personas and their interactions in more detail.

Vivienne and Linda are working on the same project, but for different companies: Vivienne is a 27-year-old technology consultant and works for 'Consulting', a large auditing and consulting firm. Linda is 25 years old and has recently started her job at Client, a company that hired Vivienne's organization for a limited period. Linda works as an assistant and is therefore responsible for booking meeting rooms, organizing team events and handling all project-related bills and invoices. Vivienne, on the other hand, is responsible for managing and assigning access rights to project-related communication platforms. She also has a deeper understanding of technology, while Linda is only familiar with tools and systems she uses every day. Both women have a similar attitude towards IT security and privacy and are concerned with keeping their company's data safe. Therefore, Linda always makes sure that all consultants sign a non-disclosure agreement, while Vivienne regularly attends IT security trainings to get informed about potential IT security threats and risks. Both women are very social and became friends very quickly. As a consultant, Vivienne has strong communication skills and is comfortable with starting conversations with strangers. Linda, on the other hand, is friendly, tends to trust her co-workers and is very forgetful.

Niko is 21 years old, studies business informatics at a university and is an intern at Consulting. Niko works for Tom, a partner at Consulting, and is responsible for preparing presentations, printing relevant handouts and uploading documents for his boss. Niko loves computer games, currently learns how to program and is very ambitious. He wants to get everything right and on time, which often stresses him out. Whenever Niko is stressed, he tends to leave his computer unlocked and forgets to shred Tom's documents that often include sensitive information. As an intern, Niko is not required to travel and works form Consulting's office. Tom is often gone and Niko gets bored easily. In that case, he socializes with other interns and loves to chat about Tom's projects.

Table 4. Developed personas with a description of their (T)asks, (S)kills, (A)ttitude towards security and (P)ersonality

	Vivienne, 27, Consultant at Consulting
(T)	Works in the field of Technology Consulting, manages relevant access rights at this project
(S)	Has a deep understanding of technology, well informed about newest IT solutions and software
(A)	Attends her company's IT security training regularly, aware of potential IT security threats, such as social engineering, tries to avoid potential security threats at all costs
(P)	Communicative and open minded, quickly became friends with Linda, often grabs a coffee at Client's kitchen to catch up with Linda
	Linda, 25, Assistant at Client
(T)	Responsible for booking meeting rooms, organizing team events and handling project-related bills and invoices
(S)	Familiar with tools she uses every day, not very familiar with any other of her company's systems
(A)	Concerned with keeping her company's data safe, ensures all consultants sign a non-disclosure agreement
(P)	Forgetful,trustworthy towards her co-workers, tells her co-workers that she cannot remember her password
	Barbara, 44, Project Lead at Consulting
(T)	Plans, coordinates and controls the project at Client, responsible for informing the sponsor of the project about its current state
(S)	Has 16+ years of experience
(A)	As a project lead, she has access to every room at Client's office, concerned with keeping any client or project-relevant data safe
(P)	Required to travel a lot, spends four days a week on a project at her client's office, works from home or at her company's office on Fridays
	Hans, 56, Head of IT at Client
(T)	Ensures Client's systems run smoothly, updates security features, checks if access rights are assigned correctly
(S)	Knows his company's systems very well
(A)	IT security has the highest priority, spends hours getting informed about potential IT risks and how they can be prevented
(P)	Passionate about his job, launched an anti-social engineering campaign at Client, informs his colleagues about adequate security behavior
	Tom, 48, Partner at Consulting
(T)	Responsible for generating revenues by acquiring new clients, makes sure existent clients are happy, supervises various clients and projects
(S)	Grew up without computers, expects his computer to work, relies on his assistant's help when it comes to fixing computer problems
(A)	Tries not to expose any sensitive information in public or while working remotely, makes an effort to use visual protection for his computer screen
(P)	Forgetful, often leaves relevant handouts behind, travels a lot due to his position
	Gabriele, 64, Project Sponsor at Client
(T)	Responsible for allocating resources efficiently, ensures projects are executed on time
(S)	Familiar with the tools she uses a lot, not very familiar with the tools she doesn't use regularly
(A)	Careful about revealing her company's information to any of the consultants
(P)	Not very trusting towards consultants, often has a hard time understanding their recommendations
	Niko, 21, Intern at Consulting
(T)	Responsible for preparing presentations, printing handouts and uploading relevant documents online
(S)	Studies business informatics, has a good understanding of IT due to his studies at a university, is learning how to program
(A)	Not aware of potential IT security threats, not very concerned with revealing sensitive data or information
(P)	New to the consulting industry, ambitious and therefore often stressed and forgetful

Tom has been with the company for more than eighteen years and is 48 years old. As a partner at Client, he is responsible for generating revenues by acquiring new projects and clients and making sure that existing clients are happy, which requires him to travel a lot. He just left his office and is currently on his way to Client. Over the last couple of years, Tom has become forgetful and started to leave printed documents behind. Tom often works remotely and always tries to get as much work done as possible. He often participates in conference calls with his colleague Barbara and employees of Client, Hans and Gabriele. Barbara is 44, has more than 16 years of professional experience and works at Consulting as a project lead. She takes her role very seriously and is responsible for planning, coordinating and reporting this project's current status to Gabriele. Barbara is concerned with keeping any client or project-relevant data safe and, like most professional consultants, spends four days a week at Client's office. On Fridays, she either works from home or her company's office.

Gabriele is 64 years old and the CFO of Client. She is responsible for allocating her company's resources efficiently and ensures that all projects are executed on time. Due to her background in finance, Gabriele knows everything about Client's financial IT tools and systems. However, she is not very familiar with any other tools at Client. She is also very cautious about revealing her company's information to any of the consultants, especially after she started working with Hans. Hans is 56 years old and Client's Head of IT. He has dedicated his life to his department and makes sure that all systems run smoothly and Client's security features are up to date at all times. Hans knows all of his company's systems very well and often checks if all access rights were assigned correctly. IT security has the highest priority for Hans, he spends hours researching potential IT threats and how they can be prevented. He has just launched an anti-social engineering campaign at Client and uses every chance to inform his colleagues about adequate security behavior.

Today, Vivienne is not required to take part in this meeting. She often works from Client's kitchen and grabs a coffee with Linda. The two have been getting along great. Linda is always excited to catch up with Vivienne, grab a cup of coffee and have a chat about work and personal matters.

5 Evaluation

In this section, we describe the evaluation process of the scenario. It was used to evaluate our methodology's outcome, since the quality of the developed scenario and personas is the main goal of the proposed method.

The evaluation sessions were structured as follows: the participants of the session were introduced to this work's main goal, the development of a consulting services scenario for HATCH, and shown a video about social engineering in order to clarify the term social engineering, its key elements and techniques. Subsequently, any emerged questions were answered and all participants were introduced to HATCH, the game's rules, scoring sheet, scenario and personas. Next, HATCH was played according to its rules, ensuring that each player at least

takes three turns. At the end of each session, all participants were first briefly asked about the game itself to prevent that a misunderstanding of the elements and rules of HATCH would influence scenario's evaluation. We then asked the players to evaluate the scenario, particularly in regards to its comprehension, completeness and closeness to reality. The provided feedback was audio recorded and subsequently analyzed.

We did not aim to evaluate HATCH's rules, game elements or mechanics and wanted to ensure that participants of the evaluation session are not distracted from the consulting services scenario. Therefore, HATCH was not elaborated any further after the participants claimed that its rules and key elements were clear and easy to understand.

In regards to HATCH's scenario, all participants agreed and stated that the represented consulting services scenario and personas are intuitive[3], easy to understand[4] and very realistic[5]. When asked for an extension of the scenario, participants suggested that the presented scenario could be extended by additional personas. While participants of the first evaluation round suggested to include an office administrator or a receptionists, members of the second session argued for adding an external service provider such as security or a cleaning personnel[6].

6 Discussion

In this section, we first discuss the results of the evaluation, followed by considerations how the presented approach can be applied in future scenarios. At the end of this section, we discuss limitations of our research.

6.1 Scenario

Reflecting the feedback of the evaluation session, it is necessary to discuss if the created consulting services scenario should be extended by additional personas, such as an office administrator, receptionist, cleaning or security personnel. On the one hand, additional personas could potentially enrich the scenario and make the serious security-awareness game more engaging and fun. On the other hand, too many personas within the scenario increase its level of complexity, make the game more difficult to play, since players need more time to go through the persona descriptions.

[3] [ES1: 1:38] "The description of the different people is very intuitive and very simply [...] modeled, also because of the figure. You could recognise it [...] very clearly".

[4] [ES2: 2:46] "Persons were described clearly and very realistic. I am able to imagine exactly how the person might be in real life, because these different types of people really exist".

[5] [ES1: 5:35] "The scenario was definitely realistic and also the [...] markers are intuitive".

[6] [ES2: 04:10] "if I am an outsider and I somehow sneak into the office, I still have to pass some [...] security guard or receptionist, that is still an upstream step, which should also be considered, I think".

Therefore, firstly we recommend including a justified and reasonable number of personas within a scenario. For example, a guard and a cleaner both represent employees of an external service provider over whom the two companies have only limited authority. Including these personas within the scenario might not contribute too much to raising employees' awareness, however will likely result in requests for establishing a security policy for externals (if not already in place). However, if they are included, it might be a reasonable trade-off to only include one or the other.

Secondly, we suggest summarizing similar roles, tasks, skill sets and attitudes towards security or privacy in one persona wherever possible. For example, receptionists and office administrators perform very similar tasks, such as handling incoming calls, arranging meetings, planning events, organizing meeting rooms and handling invoices and expenses, and therefore might resemble in their daily tasks and IT skills. However, it is also very likely that administrators/receptionists of different companies differ in their attitudes towards privacy and security. Considering all arguments, for the next version, we would extend the presented scenario by two additional personas: an administrator/receptionist who is employed by each of the respective companies, Consulting and Client.

As our study was done in 2017, we also considered the changes within the consulting industry, for example that the number of female consultants has increased [15], which is already at a reasonable level within our scenario.

6.2 Methodology

The feedback of the evaluation sessions also allows a second conclusion: the applied method for creating a scenario for a serious security-awareness game was successful, since all participants agreed that the scenario and its personas are intuitive, easy to understand and very realistic. However, since the applied method is very time-consuming and requires a lot of effort, it only makes sense under certain circumstances. One use case is, if the respective company plans to play the game on a regular basis or with a large number of players. Another use case of the derived scenario is, while being specific being generic enough to be used by other organization within the same industry (here: consulting).

6.3 Threats to Validity and Limitations

All participating interviewees were approached based on existing contacts, which could lead to a selection bias. The latter was a consequence that trials to attract 'external' consultants for interviews without payment failed, since we did not have any funding. However, the participating interviewees still had diverse properties such as position, age, gender, etc. Furthermore, it could be argued that only nine interviews were conducted. However, even within nine interviews, we could observe some satiation manifesting in a repetition of answers and similar views and statements of the experts. In the same manner, since there is no clear definition of the term 'expert' in this context, one could question our sampling. However, according to the definition of Meuser and Nagel [18], experts

are individuals who carry specific knowledge, emphasizing with the term 'specific' that the knowledge should not reflect everyday knowledge or common sense. Thus, despite experts were chosen purely based on the judgement of this work's authors, since they all work in an consulting company, they share specific knowledge about day-to-day work and processes, and therefore can be considered as experts and appropriate participants for our study.

In addition to that, it could be argued that this work's findings are not reliable, since the interview and coding process (open and axial coding), was done in two different languages: Interviews and open coding was done in German, all propositions were later summarized in English. However, we still assume that executing the interviews in the interviewees' native language is beneficial for the outcome and the translation at the end does not harm the result.

Furthermore, received answers during interviews and evaluation sessions might be subject to response biases, since we can not rule out that interviewed participants answered what they assumed the interviewer wants to hear or is socially acceptable. We tried to address that by not using any triggering terms and did not push for a response, allowing the interviewees a way out by not answering the questions.

6.4 Future Work

We suggest further validation of our method and its results to investigate if it can be transferred to another organization or domain. Additionally, we suggest to investigate if in the same manner or with which changes, a scenario and personas could be derived for a similar serious games on social engineering.

Additionally, we think that as future work it should be evaluated if the effort can be reduced, for example by conducting less or shorter expert interviews. In addition to that, we believe that the process of deriving an interview guide can be shortened and based on the interview guide presented in this paper, since all questions are directed towards the game's key elements, which are the industry's specifics, assets, communication channels, location and existing personas.

Hill et al. [14] showed that the use of multiple photos (of males and females) for a single persona to avoid gender stereotypes did not reduce project designers' engagement with the personas. Thus, another interesting question, far beyond the scope of this work, is if the use of multiple photos for a single persona would change players' engagement with HATCH's personas.

7 Conclusion

In this paper, we added to addressing the problem that many firms do not address social engineering security threats adequately or only apply ineffective defense mechanisms, such as traditional trainings, penetration tests or standardized security awareness campaigns or serious games. We proposed to create specific scenarios considering the the organisation's specifics and based on the work of Faily and Flechais [11] proposed a method to develop a new scenario for HATCH.

The result of our research is that our method for adapting a serious game on social engineering was effective, since all participants of the evaluation sessions agreed that the derived scenario and its personas are realistic. However, the proposed method is also very time-consuming, requires a lot of effort and only makes sense if the scenario can be used several times by an organization or can be transferred to another, similar organization. We propose future work to investigate if the effort can be reduced.

Acknowledgements. This work was supported by European Union's Horizon 2020 research and innovation program from the project CyberSec4Europe (grant agreement number: 830929) and from the project THREAT-ARREST (grant agreement number: 786890). We are grateful for image services of Pixabay, Pexels and Unsplash.

References

1. Abawajy, J.: User preference of cyber security awareness delivery methods. Behav. Inf. Technol. **33**(3), 237–248 (2014)
2. Alexander, M.: Methods for understanding and reducing social engineering attacks. SANS Inst. **1**, 1–32 (2016). https://www.sans.org/reading-room/whitepapers/critical/methods-understand-ing-reducing-social-engineering-attacks-36972
3. Bada, M., Sasse, A.M., Nurse, J.R.C.: Cyber security awareness campaigns: Why do they fail to change behaviour? CoRR abs/1901.02672 (2019). http://arxiv.org/abs/1901.02672
4. Beckers, K., Pape, S.: A serious game for eliciting social engineering security requirements. In: Proceedings of the 24th IEEE International Conference on Requirements Engineering, RE 2016. IEEE Computer Society (2016)
5. Beckers, K., Pape, S., Fries, V.: HATCH: Hack and trick capricious humans - a serious game on social engineering. In: Proceedings of the 2016 British HCI Conference, Bournemouth, United Kingdom, 11–15 July 2016 (2016)
6. Charmaz, K.: Constructing Grounded Theory. Sage, Thousand Oaks (2014)
7. Cooper, A.: The inmates are running the asylum. indianapolis, ia: Sams. Macmillan (1999)
8. Corbin, J., Strauss, A.: Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory. Sage publications, Thousand Oaks (2014)
9. Dimkov, T., Van Cleeff, A., Pieters, W., Hartel, P.: Two methodologies for physical penetration testing using social engineering. In: Proceedings of the 26th Annual Computer Security Applications Conference, pp. 399–408 (2010)
10. Donovan, L., Lead, P.: The use of serious games in the corporate sector. A State of the Art Report. Learnovate Centre (2012)
11. Faily, S., Flechais, I.: Persona cases: a technique for grounding personas. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 2267–2270 (2011)
12. Flick, U.: An Introduction to Qualitative Research. Sage, Thousand Oaks (2014)
13. Gläser, J., Laudel, G.: Experteninterviews und qualitative Inhaltsanalyse: als Instrumente rekonstruierender Untersuchungen. Springer, Heidelberg (2009)
14. Hill, C.G., et al.: Gender-inclusiveness personas vs. stereotyping: can we have it both ways? In: Proceedings of the 2017 Chi Conference on Human Factors in Computing Systems, pp. 6658–6671 (2017)

15. Huang, J., Krivkovich, A., Starikova, I., Yee, L., Zanoschi, D.: Women in the workplace 2019. McKinsey & Company and LeanIn.Org (2019). https://www.mckinsey.com/~/media/McKinsey/Featured%20Insights/Gender%20Equality/Women%20in%20the%20Workplace%202019/Women-in-the-workplace-2019.pdf
16. Kipker, D.K., Pape, S., Wojak, S., Beckers, K.: Juristische bewertung eines social-engineering-abwehr trainings. In: Rudel, S., Lechner, U. (eds.) State of the Art: IT-Sicherheit für Kritische Infrastrukturen, pp. 112–115. Universität der Bundeswehr, Neubiberg (2018). https://www.itskritis.de/_uploads/user/IT-Sicherheit%20Kritische%20Infrastrukturen%E2%80%93screen.pdf#page=112
17. Malheiros, M., Jennett, C., Seager, W., Sasse, M.A.: Trusting to learn: trust and privacy issues in serious games. In: McCune, J.M., Balacheff, B., Perrig, A., Sadeghi, A.-R., Sasse, A., Beres, Y. (eds.) Trust 2011. LNCS, vol. 6740, pp. 116–130. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-21599-5_9
18. Meuser, M., Nagel, U.: The expert interview and changes in knowledge production. In: Bogner, A., Littig, B., Menz, W. (eds.) Interviewing Experts, pp. 17–42. Springer, Heidelberg (2009). https://doi.org/10.1057/9780230244276_2
19. Mitnick, K.D., Simon, W.L.: The Art of Deception: Controlling the Human Element of Security. John Wiley & Sons, Hoboken (2003)
20. Naderer, G., Balzer, E., Batinic, B., Bauer, F., Blank, R., David, J.: Qualitative Marktforschung in Theorie und Praxis. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-8349-6790-9
21. Papadaki, M., Furnell, S., Dodge, R.: Social engineering: Exploiting the weakest links. European Network & Information Security Agency (ENISA), Heraklion, Crete (2008)
22. Peltier, T.R.: Social engineering: concepts and solutions. Inf. Secur. J. **15**(5), 13 (2006)
23. Petridis, P., et al.: State of the art in business games. Int. J. Serious Games **2**(1) (2015)
24. Pruitt, J., Adlin, T.: The Persona Lifecycle: Keeping People in Mind Throughout Product Design. Elsevier, Amsterdam (2010)
25. Riedel, J.C., Hauge, J.B.: State of the art of serious games for business and industry. In: 2011 17th International Conference on Concurrent Enterprising, pp. 1–8. IEEE (2011)
26. Schaab, P., Beckers, K., Pape, S.: A systematic gap analysis of social engineering defence mechanisms considering social psychology. In: Proceedings of the 10th International Symposium on Human Aspects of Information Security & Assurance, HAISA 2016, Frankfurt, Germany, 19–21 July 2016 (2016)
27. Schaab, P., Beckers, K., Pape, S.: Social engineering defence mechanisms and counteracting training strategies. Inf. Comput. Secur. **25**(2), 206–222 (2017)

A.9 Conceptualization of a CyberSecurity Awareness Quiz

Sebastian Pape, Ludger Goeke, Alejandro Quintanar, and Kristian Beckers. Conceptualization of a cybersecurity awareness quiz. In *Computer Security - ESORICS 2020 International Workshops MSTEC*, volume 12512 of *LNCS*, pages 61–76, Cham, 09 2020. Springer International Publishing. doi: 10.1007/978-3-030-62433-0_4. URL https://link.springer.com/chapter/10.1007%2F978-3-030-62433-0_4

Conceptualization of a CyberSecurity Awareness Quiz

Sebastian Pape[1,2(✉)], Ludger Goeke[1], Alejandro Quintanar[1], and Kristian Beckers[1]

[1] Social Engineering Academy (SEA) GmbH,
Eschersheimer Landstrasse 42, 60322 Frankfurt am Main, Germany
Sebastian.Pape@m-chair.de
[2] Faculty of Economics and Business Administration, Goethe University Frankfurt,
Theodor-W.-Adorno-Platz 4, 60323 Frankfurt am Main, Germany

Abstract. Recent approaches to raise security awareness have improved a lot in terms of user-friendliness and user engagement. However, since social engineering attacks on employees are evolving fast, new variants arise very rapidly. To deal with recent changes, our serious game *CyberSecurity Awareness Quiz* provides a quiz on recent variants to make employees aware of new attacks or attack variants in an entertaining way. While the gameplay of a quiz is more or less generic, the core of our contribution is a concept to create questions and answers based on current affairs and attacks observed in the wild.

Keywords: Serious game · CyberSecurity Awareness · Human factor

1 Introduction

Social engineering attacks represent a continuing threat to employees of organizations. With a wide availability of different tools and information sources [5], it is a challenging task to keep up to date of recent attacks on employees since new attacks are being developed and modifications of known attack scenarios are emerging. The latest Data Breach Investigations Report [2] reports another increase of financially motivated social engineering, where the attacker directly ask for some money, i. e. by impersonating CEOs or other high-level executives. However, during the writing of the report, scammers have already varied their approach and also ask for purchase and transfer of online gift cards[1] in order to scam employees. Additionally, scammers also base attacks on the current news situation, such as COVID-19 Ransomware [15]. While a couple of defense methods and counteracting training methods [16,17] exist, at present, most of them can not be adapted fast enough to cope with this amount and speed of new variations.

[1] https://twitter.com/sjmurdoch/status/1217449265112535040.

G. Hatzivasilis and S. Ioannidis (Eds.): MSTEC 2020, LNCS 12512, pp. 61–76, 2020.
https://doi.org/10.1007/978-3-030-62433-0_4

The *CyberSecurity Awareness Quiz* is a serious game in form of an online quiz to raise the security awareness of employees, in particular against social engineering attacks. The game follows the approach that quiz questions are based on real-world social engineering attacks. Additionally, the pool of questions will constantly be extended by new questions in relation to current social engineering attacks. For this purpose, a specific process for the procurement of appropriate information is developed, which is described in detail in Sect. 3.2. Our contribution within this paper is the conceptualization of the *CyberSecurity Awareness Quiz* with a focus on the concept how to generate questions for the quiz game based on current affairs and attacks observed in the wild.

The remainder of the paper is structured as follows: Sect. 2 lists some related games, explains the relationship of the *CyberSecurity Awareness Quiz* with previously developed games and how it integrates into a more general training platform. Its concept is explained in Sect. 3 along with the planned components in Sect. 4. We conclude in Sect. 5.

2 Background and Related Work

There is a large number of tabletop games for security training or awareness raising [3,4,6,8,14] targeting different domains, asset and areas in the academia.

However, the ones which are closer to *CyberSecurity Awareness Quiz* are mostly commercial without a detailed description. Nevertheless, we give a brief overview of them in the following. The "Emergynt Risk Deck" highlights IT-security risks to business leadership [7]. "OWASP Snakes and Ladders" is an educational game to raise security awareness about application security controls and risks [13]. Within the game "Quer durch die Sicherheit" players move towards the target by answering questions correctly [10]. "Stadt Land HACK!" is a quiz about data privacy and security [11].

Since the above mentioned games are all tabletop or card games, they can not be adapted to recent security incidents easily. While there is only a limited variation of different variants of a quiz-style game, our main contribution of this conceptual paper is the process for the creation of questions along with the idea to mostly use the *CyberSecurity Awareness Quiz* to keep users informed about recent attacks in an entertaining way.

2.1 Relation to Existing Games

Naturally, the aim and scope of a game can not be too broad. Similar to security awareness campaigns [1], serious games also benefit from an adaption to the user and his/her specific needs. Therefore, *CyberSecurity Awareness Quiz* is part of a series of games dovetailed to a chain aiming at raising security awareness (cf. Fig. 1). For security requirements engineering, employees are playing HATCH [3], in order to identify relevant attacks and develop countermeasures. All identified threats which can not be technically addressed, need to be integrated into the organisation's security policy. Once the security policy is developed or updated,

employees can train to apply it and get an understanding how it addresses certain attacks by playing PROTECT [9]. However, naturally different attacks or variations of attacks will sprout faster than the security policies can be adapted. Thus, *CyberSecurity Awareness Quiz* is used to raise awareness about the latest attacks and their variations, based on the player's general understanding developed in the game sessions of HATCH and PROTECT.

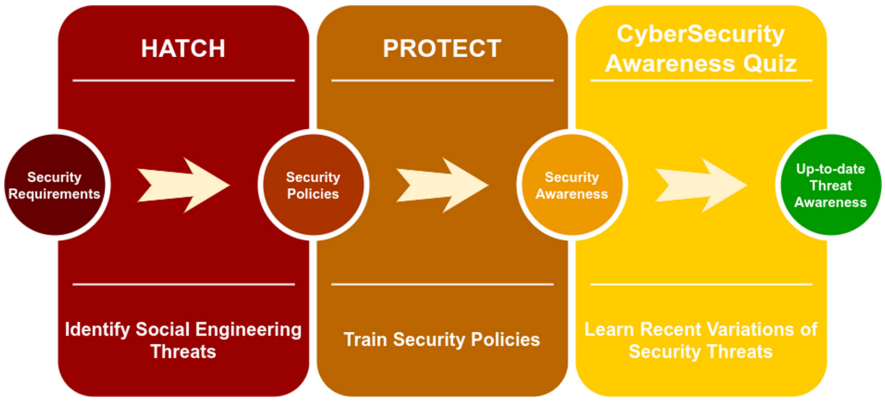


Fig. 1. The relation of HATCH [3], PROTECT [9] and *CyberSecurity Awareness Quiz*

2.2 Embedding into a CyberSecurity Training Platform

Besides the use and interplay of *CyberSecurity Awareness Quiz* with other serious games, it is also important to integrate them into a more general training platform, such as the THREAT-ARREST [12] advanced training platform (cf. Fig. 2).

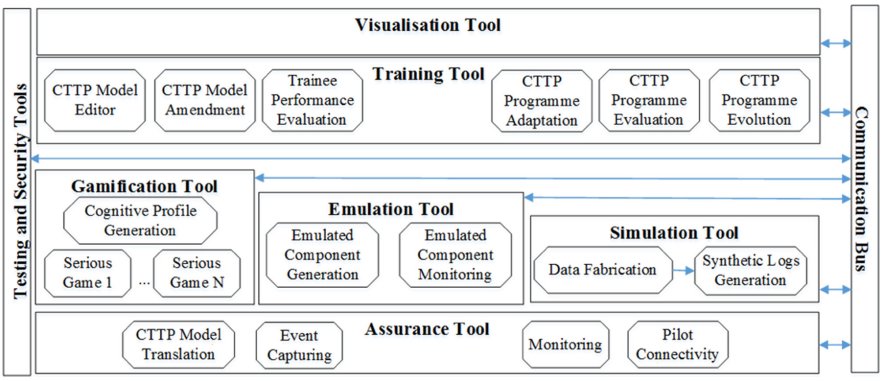


Fig. 2. The THREAT-ARREST advanced training platform [12]

This way it is not only possible to train employees during their use of the serious games, but also to embed and manage their efforts in a broader way.

The result of *CyberSecurity Awareness Quiz* sessions contribute to THREAT-ARREST's continuous evaluation of the individual trainees' performance and the effectiveness of the training programs. Within the platform for each trainee results of the serious games, the emulation, the simulation and the training tool are brought together to spot possible gaps in the employee's knowledge or awareness. If knowledge gaps are identified, it can be checked if there already exists a training on the specific topic as serious game, simulation or emulation of the cyber range system. If no appropriate training can be identified, this might indicate the need of producing a new training, tailored to the organizational needs and the trainee types.

3 Concept

The fast change and adaption of attacks as sketched in the introduction show the necessity for employees to keep their knowledge about social engineering up-to-date.

Since we expect only a reasonable amount of new attacks or attack variations, we decided to aim for a lightweight game with the idea that it could be played occasionally (e.g. when traveling in trams or subways). In general, the game should be playable alone since this avoids any necessity to find or wait for other players, but in particular for long term motivation, comparisons with or games against other players should be possible. In summary, we identified the following requirements:

- Questions refer to recent real-world threats
- Lightweight
- Playable on mobile devices
- Single and multi-player modes

3.1 Game Concept

One game type which fulfills the requirements is a quiz game, where players have to answer a set of questions. In *CyberSecurity Awareness Quiz*[,] a question describes a certain social engineering attack scenario which is based on a recent attack observed in the real world in an abstract and general way. For every question, the possible answers contain one or more correct answers and one or more incorrect answers. Correct answers will represent consequences which result from the attack that is described in the question. Accordingly, incorrect answers will represent effects which can not result from the attack. A mockup of the planned GUI which also shows a sample question is illustrated in Fig. 3.

CyberSecurity Awareness Quiz will provide different modes in which a quiz can be played. Either by a single player or in competition between two players. These modes are described in the following:

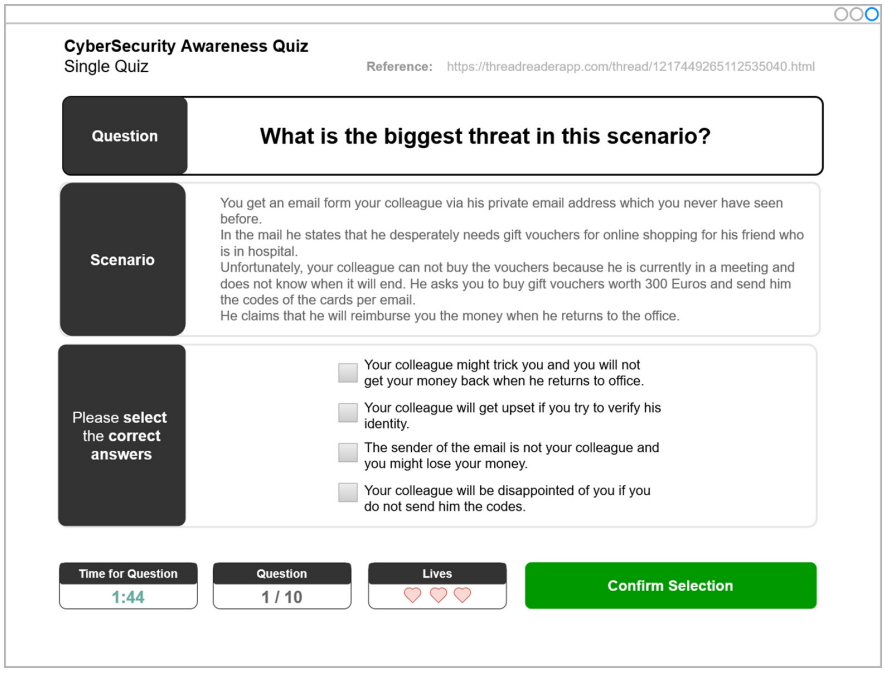


Fig. 3. Mockup of the user interface along with a sample question

Single Quiz: A player will answer the questions of a quiz alone.

Context Quiz: Single-player quiz with specific questions depending on the preferences of a player. Examples for specific questions are scenarios concerning a certain location, industry sector or role/position in the company. Furthermore, it is possible to play only recent added questions, e.g. questions added in the last 3 months.

Versus Quiz: Two players will compete in a quiz against each other. A question will be asked simultaneously to both players. The player who will answer a question correctly gets a point. If both players are correct, the faster player wins. The player who will answer more questions correctly, wins the quiz round.

Pick Quiz: In this mode, two players will answer questions one after the other. Here, the player who has answered his/her last question correctly chooses the next question for the opponent out of different options until the opponent answers a question correctly. If this is the case, the right for choosing questions changes and so on. Only the first question will be asked to both players simultaneously. The player who answers this question correctly first will have the right to choose the next question for the opponent.

Draw Quiz: This mode will have the same rules as the *Pick Quiz* mode with the following modification: Instead of choosing the next question out of

different options, the player who has answered his/her last question correctly will choose the industry/sector to which the next question for the opponent relates.

For the modes context quiz, pick quiz and draw quiz, certain metadata on the scenarios is needed. Therefore, question will be tagged by predefined types of metadata. This metadata will enable a categorization of questions which allows it to combine questions to different quizzes for certain training objectives or specific groups of players. For example, a specific set of questions will be able to reference a certain type of attacks (e.g. different forms of phishing), industry sector (e.g. energy suppliers), department (e.g. human resources), a geographic area (e .g. Europe) or all new attacks added after a given date. The possibility of adapting a quiz to the players needs aims to enable players to map the mediated learning content directly to their work routine.

Additionally, the metadata will enable an on the fly compilation of the questions for a quiz round played in the Context Quiz mode. Here, the player provides information which refers to certain aspects of social engineering he/she wants to be considered in the next quiz round. This quiz round will include all the predefined questions which are tagged with metadata that matches the provided information.

We describe the different types of metadata used in Sect. 3.2.

3.2 Process for Information Procurement and Question Generation

A key feature of *CyberSecurity Awareness Quiz* will be the fact that its questions are based on real-life attacks whereby the amount of questions will be permanently expended to cover new social engineering attacks. To fulfill this requirement, an appropriate process for gathering content regarding attacks and the creation of corresponding questions and answers is needed. This process is sketched in Fig. 4.

The first step of the process includes the procurement of information with respect to current social engineering attacks. While the number of relevant attacks might be feasible, there is a huge amount of reports of attacks, privacy breaches, data losses, etc. Due to the high frequency in which they occur as well as the multitude of information sources, the information procurement presents an enormous challenge. To meet this challenge, the information procurement will include automated tasks which are discussed later in this section.

The second step of the process for the creation of questions and answers includes the formulation of questions for a quiz. Usually, questions will be created based on content about social engineering attacks which has been collected in Step 1. If this is the case, the game content designer will check for a new relevant web feed first if a corresponding question already exists. For this check he/she will filter the existing questions by the types of metadata which are relevant for the new web feed.

In the third step, a created question will be tagged with metadata. This metadata will represent characteristics of an attack like the category of an attack

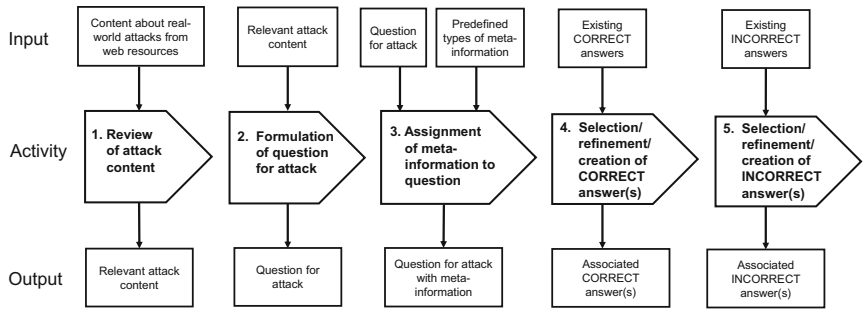


Fig. 4. Process for the creation of questions and answers for social engineering attacks

(e.g. phishing). *CyberSecurity Awareness Quiz* will provide predefined types of metadata, which are specified in Table 1. This table includes the name of a metadata type and its description. The metadata of questions is important for the reuse of questions during the creation of certain predefined quizzes and the compilation of on the fly quizzes within the Context Quiz mode (see Sect. 3.1). As discussed in the previous section, metadata allows to filter questions by special categories when creating a quiz with a certain topic. For example, if a quiz shall refer to attacks which are targeting employees of the human resources department, questions whose metadata parameter of the type *Department* has the value "human resources" should be assessed for consideration. The same concept is applied when a quiz round is played in the Context Quiz mode. Here, the player provides information regarding his/her preferences and the started quiz comprises only such questions whose metadata corresponds to the provided information. For example, if a player is interested in all types of new phishing attacks from a certain point in time, he/she can selects the value "phishing" for the metadata type *Attack category* and the value "from 01.06.2020" for the metadata type *Time of attack*.

In the fourth step of the process **correct** answers are assigned to a question. In this context, new correct answers can be created or already existing correct answers can be reused.

The last step of the process includes the assignment of **incorrect** answers to a question. As for correct answers, incorrect answers can be newly created or already existing incorrect answers can be reused.

Information Procurement. One objective of the information procurement is to gather content related to social engineering attacks which is published on appropriate web resources like news websites, websites about information security, websites of institutions, blogs or even twitter. In this context, in particular websites which provide information about their new content in a structured manner (e.g. web feeds) will be considered. Figure 5 shows an overview of the steps for the information procurement.

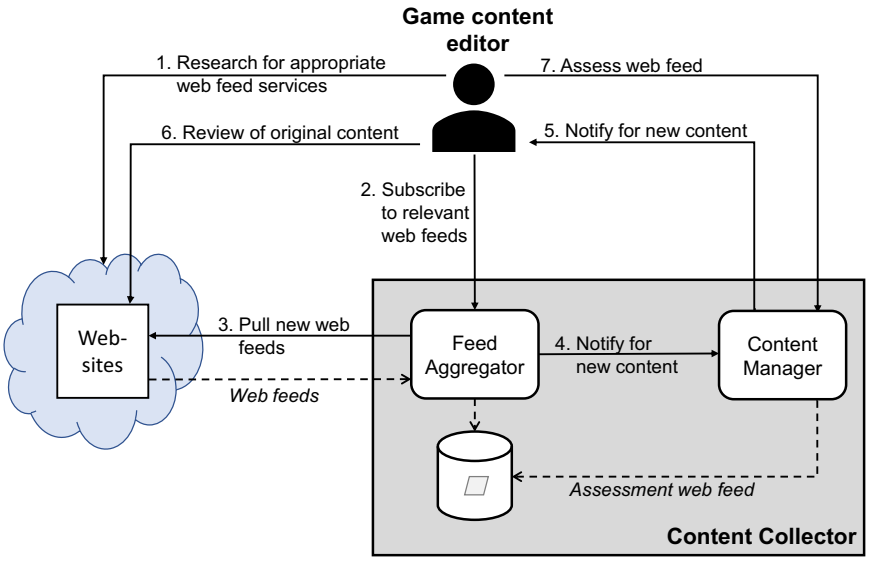


Fig. 5. Tasks for gathering and analysing content about attacks

Web feeds present a form of pull data. This means, that users can request frequently information in relation to new content on subscribed websites by using appropriate tools (e.g. feedreaders). Web feeds are machine-readable files which are provided in standardized formats like RSS[2] or Atom[3]. They include data which addresses among others the title and a short description of the new content, the URL of the original resource, the publishing date and the name of the author.

As Fig. 5 illustrates, some tasks for the information procurement need to be performed manually by the *game content editor*. Other steps will be performed automatically by a component of *CyberSecurity Awareness Quiz* which is named *Content Collector*. The different steps of the process for information procurement are explained in the following.

In the initial step of the process, the game content editor will search for websites which publish content about social engineering attacks and implement a web feed service. This step will be repeated periodically to check if new appropriate web resources are available. In the second step, the game content editor will subscribe to the found web feed services by using the *Feed Aggregator* which is a subcomponent of the Content Collector. The Feed Aggregator will query automatically and periodically the subscribed websites for new web feeds (step 3). If new web feeds have been found, it will notify the *Content Manager*

[2] depending on the version RSS means: RDF Site Summary or Really Simple Syndication.

[3] Atom Syndication Format is an XML language used for web feeds.

(step 4) which is another subcomponent of the Content Collector. The Content Manager, which is responsible for the management of gathered web feeds, will inform the game content editor that new content is available (step 5). Then, the game content editor will review the original content of the corresponding web feed (step 6). Afterwards he/she will assess in the Content Manager if the content to the web feed is relevant or not (step 7).

Web feeds which will be marked as relevant can be used for the formulation of new quiz questions (see Fig. 4, step 2).

Types of Metadata. As already discussed, questions need to be tagged by metadata in order to allow the categorization of questions during the creation of predefined quizzes and within on the fly compilation of quizzes with respect to the Context Quiz mode (see Sect. 3.1). The different types of metadata are specified in Table 1. Additionally, (correct and incorrect) answers will be also tagged with metadata (cf. Table 2). The *Multiplicity* will specify the number of data items which have to be assigned at least and can be assigned at most.

Table 1. Types of metadata for tagging of questions

Type of metadata	Description	Multiplicity
Title	Title of an attack	1
Type of attack execution	Specification if an attack is executed (i) directly on site by an attacker (e.g. an attacker tries to get access to a secured server room by pretending to be a service technician), (ii) indirectly by using a technical medium (e.g. phishing via email) or (iii) different combinations of direct and/or indirect executions	1
Attack category	Categories which typify an attack (e.g. vishing). In this connection, an attack can be assigned to exactly one category or to several categories. For example, an attack which uses dumpster diving can only be associated to the category *dumpster diving*. An attack in which emails with malicious links are sent to CEOs can be assigned to the categories *email fraud*, *phishing*, *email phishing* and *whaling*	1..*
Type of attacker	Typing of the attacker who executes an attack (e.g. cyber criminal, fraudster, intelligence service, hacker)	1..*
Feigned identity	Defines the identity of the entity/person which/who is feigned by the attacker during an attack. Regarding enterprises or institutions, a feigned identity could refer to internal persons like colleagues, C-level personnel and employees from other branches or external persons like customers, technicians and cleaning stuff. In the private context, an attacker could pretend to be a relative, friend or a person who seeks for help. When feigning an entity, an attacker could pretend to be an employee of a state authority (e.g. tax authority) or a private institute (e.g. banks)	1..*
Context of victims	Specifies the context(s) of the victims who are targeted by an attack. For this parameter the values *individual* and *organisation* are predefined	1..2
Characteristics of private victim	Specifies the characteristic(s) for a group of victims in person of *individual* who are threatened by an attack. For individuals this could be demographic characteristics (e.g. age, gender, interests, internet usage)	0..*

(*continued*)

Table 1. (*continued*)

Type of metadata	Description	Multiplicity
Sector	Specifies the sector/industry of organisations which are threatened by an attack (e.g. energy suppliers, financial institutes, state institutions)	0..*
Department	Defines certain departments of an organisation (e.g. human resources, finance, IT) which are affected by an attack	0..*
Role	Indicates certain roles of employees of an organisation (e.g. CEO, administrator, financial accountant) which are threatened by an attack	0..*
Motivation for attack	Specifies the motivation for the execution of an attack (e.g. espionage, criminal intend, interest in hacking)	1..*
Objective description	Defines the objective of an attack (e.g. illegal financial transactions, gaining of sensitive information/data, identity theft)	1..*
Exploited psychological pattern	Psychological pattern which is tried to be exploited by an attack (e.g. authority, good faith, laziness)	1..*
Used technology	Technology which has been used during the attack (e.g. email for phishing or telephone for vishing)	0..*
Geographical spreading	The geographical area where the attack has been conducted (e.g. worldwide, Europe, United States, California, Milan)	1..*
Time of attack	Period(s) of time in which the attack has been conducted	1..*
Sources	Sources of the content on which the attack bases	1..*

CONDITION: This parameter is only used when the parameter *Context of victims* has the value *individual*
CONDITION: This parameter is only used when the parameter *Context of victims* has the value *organisation*

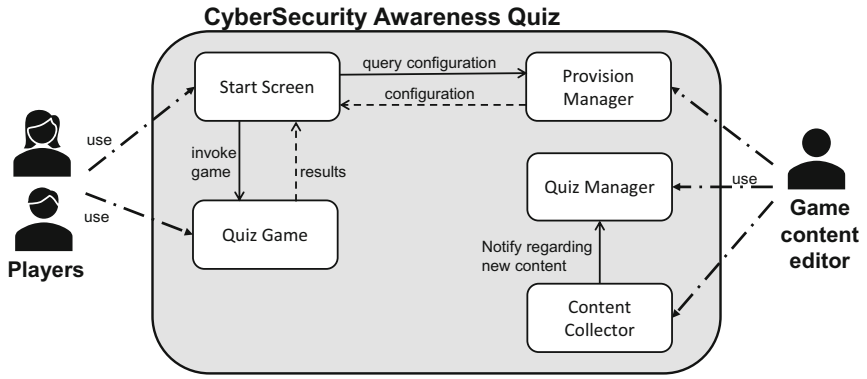


Fig. 6. Components of *CyberSecurity Awareness Quiz*

Table 2. Types of metadata for tagging of answers

Type of metadata	Description	Multiplicity
Attack category	Specifies the attack category or rather different attack categories of questions to which an answer could be assigned	1..*
Answer type	Indicates if an answer is correct or incorrect in the context of its attack categories	1

4 Architecture and Components

This section discusses the different components of *CyberSecurity Awareness Quiz* which will implement the concepts described in Sect. 3. Figure 6 provides an overview of these components and the rudimentary communication between them. Additional, it shows the different roles which will use certain components. For the sake of clarity, a representation of the database and the corresponding communication between the database and components has been omitted.

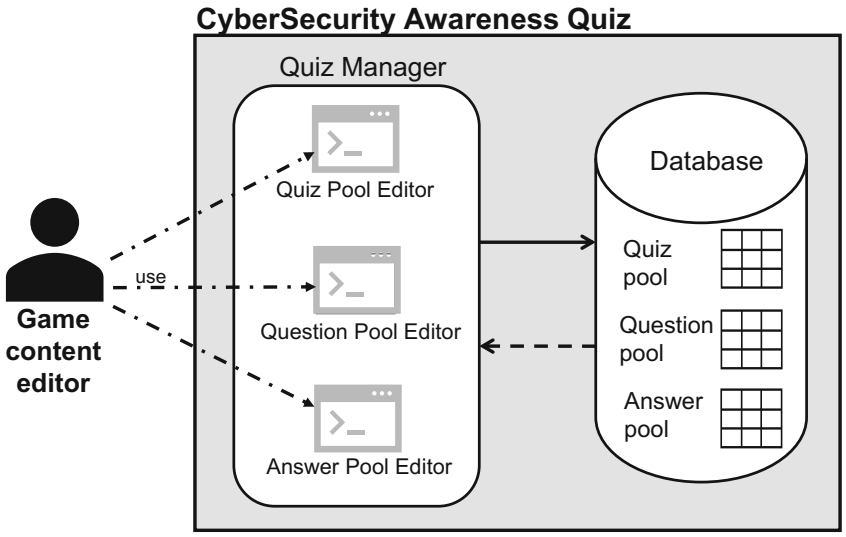


Fig. 7. Editors provided by the quiz manager

4.1 Content Collector

We have already introduced the *Content Collector* in Sect. 3.2, thus the following description is limited to the essentials.

The Content Collector will provide functionality for the collection of new content about social engineering attacks in the form web feeds. To this, it will check the subscribed web feed services frequently for new content.

A further functionality of the Content Collector will enable the management of collected web feeds. It will inform the game content designer when new content has been collected and will allow to assign his/her assessments regarding its relevance to the related web feeds. If a web feed will be considered as relevant by the game content editor, the Content Collector will notify another component in form of the *Quiz Manager* (see Sect. 4.2) that new relevant content is available.

The content collector will be exclusively used by the game content editor.

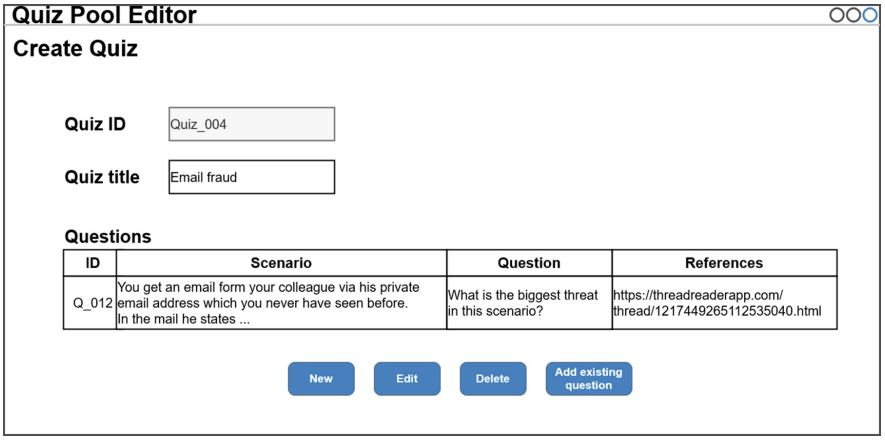


Fig. 8. Mockup of the user interface of the quiz pool editor

4.2 Quiz Manager

The *Quiz Manager* will enable the game content editor to manage (i) the pool of available quizzes and the separate (ii) pool of questions and (iii) pool of answers. For that purpose, the Quiz Manager will implement corresponding editors named *Quiz Pool Editor*, *Question Pool Editor* and *Answer Pool Editor*. These different editors, which are represented in Fig. 7, are discussed in the following.

The *Question Pool Editor* will enable the creation of questions which are added to the question pool (cf. Fig. 7) and the specification of the corresponding metadata. In general, the questions are based on content that has been collected by the *Content Collector* (see Sect. 4.1). Additionally, the Question Pool Editor will allow the editing of questions in the pool and their deletion.

In the context of creating or editing a question, the Question Pool Editor will also implement the assignment of correct and incorrect answers to a question. For that purpose, it will supply a dialogue for the creation of new answers and the related metadata. When the input is finalized, a created answer will be added to the answer pool (cf. Fig. 7).

The Question Pool Editor will also display a list of existing answers from the answer pool which could be relevant for the current question because of their assigned attack categories. Besides adding new answers, it will be possible to assign any existing answer to the edited question.

With respect to the management of the answer pool (cf. Fig. 7), the *Answer Pool Editor* will implement the creation of new answers and the related metadata as well as the editing and deletion of answers.

The functionality of creating new quizzes and adding them to the pool of available quizzes (cf. Fig. 7) will be implemented by the *Quiz Pool Editor* (cf. Fig. 8). In the mockup of the user interface of the Quiz Pool Editor it is shown that every quiz has a title and is identified by an unique identifier.

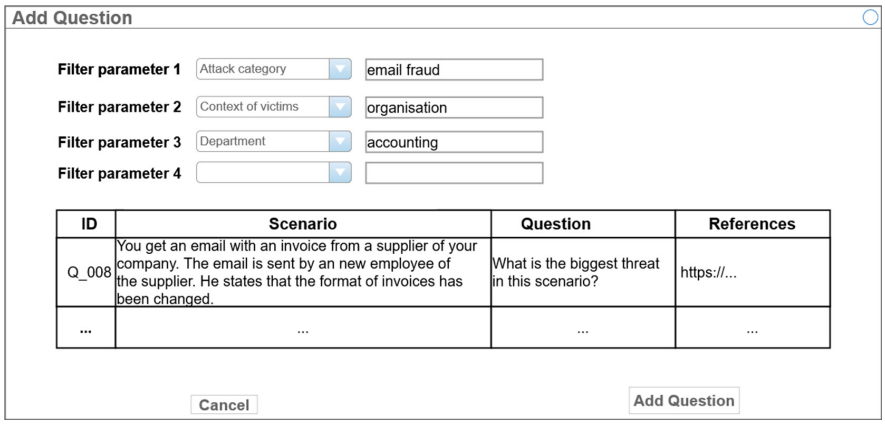


Fig. 9. Mockup of the user interface of the add question dialogue of the quiz pool manager

It will be possible to reuse predefined questions from the question pool for a new quiz. For that purpose, the Quiz Pool Editor will display a list of predefined questions from the question pool which can be filtered by the metadata of the questions. This way, the game content designer will be able to restrict the number of displayed questions.

During the creation of a quiz, the Quiz Pool Editor will also allow the creation of new questions and the related answers. A newly created question will be added additionally to the question pool, when it is finalized. If newly created answers will be assigned to a created question, these answers will be also added to the answer pool. Figure 9 shows the dialogue for adding existing questions to a quiz. Here, the set of displayed questions corresponds to the selected filter parameters.

Functionalities for the editing and deletion of quizzes will also be supplied by the Quiz Pool Editor.

4.3 Provision Manager

The *Provision Manager* facilitates configurations with respect to provisions of *CyberSecurity Awareness Quiz* These configurations will be managed by the game content editor. The different configuration parameters are represented in Table 3.

4.4 Start Screen

When a player will start the *CyberSecurity Awareness Quiz* client, the *Start Screen* will appear. Depending on the configuration provided by the *Provison Manager*, the Start Screen will show which gaming modes are activated and which quizzes can be played.

Table 3. Configuration parameters for the provisioning of *CyberSecurity Awareness Quiz*

Configuration parameter	Description
Available quizzes	Specifies the quizzes which shall be available for the player to be played
Activated modes	Indicates which single-player modes and/or multi-player modes shall be activated within a provision

The Start Screen acts as a frontend of *CyberSecurity Awareness Quiz* to start games in the component *Quiz Game* (see Sect. 4.5) with one of the activated quizzes. If the player plays a game in the *Context Quiz* mode (see Sect. 3), he/she will be able to provide the information which determines how the content of the quiz to be played will be compiled.

If any multi-player mode is activated, the Start Screen will display other players which are currently online. Accordingly, a player will be able to arrange a game in one of the multi-player modes with an available competitor.

4.5 Quiz Game

The component *Quiz Game* will implement the actual quiz game. A certain quiz game can be invoked by the *Start Screen* (see Sect. 4.4). For that purpose, the Start Screen will pass the required parameters for a quiz to the Quiz Game. These parameters will include among other information, the set of questions and the mode in which the quiz will be played.

The graphical user interface (GUI) of the Quiz Game will differ depending on the gaming mode in which the quiz is played. A mockup was already presented in Fig. 3 in Sect. 3.1.

5 Conclusion

We presented a conceptualization of *CyberSecurity Awareness Quiz* based on the requirements defined in Sect. 3. From a conceptual perspective, all requirements are fulfilled. In particular, one of our contributions is a detailed description of the process for information procurement and deduction of questions based on recent social engineering attacks. The game offers different quiz modes to maintain the players' long-term motivation and interest to gather knowledge on new attacks. Besides the obvious implementation of *CyberSecurity Awareness Quiz*[,] in future work we intend to investigate by user studies if the implementation is also perceived as lightweight by the players and if players perceive the game suitable for occasional playing.

Acknowledgements. This work was supported by European Union's Horizon 2020 research and innovation program from the project THREAT-ARREST (grant agreement number: 786890) and CyberSec4Europe (grant agreement number: 830929).

References

All URLs haven been last accessed on July 22nd, 2020.

1. Bada, M., Sasse, A.M., Nurse, J.R.C.: Cyber security awareness campaigns: why do they fail to change behaviour? CoRR abs/1901.02672 (2019). http://arxiv.org/abs/1901.02672
2. Bassett, G., Hylender, C.D., Langlois, P., Pinto, A., Widup, S.: Data breach investigations report (2020). https://enterprise.verizon.com/resources/reports/2020-data-breach-investigations-report.pdf
3. Beckers, K., Pape, S.: A serious game for eliciting social engineering security requirements. In: Proceedings of the 24th IEEE International Conference on Requirements Engineering, RE 2016. IEEE Computer Society (2016). https://doi.org/10.1109/RE.2016.39
4. Beckers, K., Pape, S., Fries, V.: HATCH: hack and trick capricious humans - a serious game on social engineering. In: Proceedings of the 2016 British HCI Conference, Bournemouth, United Kingdom, 11–15 July 2016 (2016). https://ewic.bcs.org/content/ConWebDoc/56973
5. Beckers, K., Schosser, D., Pape, S., Schaab, P.: A structured comparison of social engineering intelligence gathering tools. In: Lopez, J., Fischer-Hübner, S., Lambrinoudakis, C. (eds.) TrustBus 2017. LNCS, vol. 10442, pp. 232–246. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-64483-7_15
6. Denning, T., Lerner, A., Shostack, A., Kohno, T.: Control-alt-hack: the design and evaluation of a card game for computer security awareness and education. In: Proceedings of the 2013 ACM SIGSAC Conference on Computer & Communications Security, pp. 915–928 (2013)
7. Emergent Network Defense: Emergynt risk homepage. https://emergynt.com/risk-deck/
8. Frey, S., Rashid, A., Anthonysamy, P., Pinto-Albuquerque, M., Naqvi, S.A.: The good, the bad and the ugly: a study of security decisions in a cyber-physical systems game. IEEE Trans. Software Eng. **45**(5), 521–536 (2017)
9. Goeke, L., Quintanar, A., Beckers, K., Pape, S.: PROTECT – an easy configurable serious game to train employees against social engineering attacks. In: Fournaris, A.P., et al. (eds.) IOSEC/MSTEC/FINSEC -2019. LNCS, vol. 11981, pp. 156–171. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-42051-2_11
10. Known Sense: Quer durch die Sicherheit game reference. http://www.known-sense.de/quer_durch_die_sicherheit_folder.pdf
11. Known Sense: Stadt Land HACK! homepage. http://www.known-sense.de/stadt_land_hack.pdf
12. Koshutanski, H., et al.: Threat-arrest platform's initial reference architecture. Technical report, Threat-Arrest, Deliverable 1.3 (2019)
13. OWASP: Owasp snakes and ladders homepage (2013). https://owasp.org/www-project-snakes-and-ladders/
14. Rieb, A., Lechner, U.: Operation digital chameleon: towards an open cybersecurity method. In: Proceedings of the 12th International Symposium on Open Collaboration, pp. 1–10 (2016)

15. Saleh, T.: Covidlock update: deeper analysis of coronavirus android ransomware (2020). https://www.domaintools.com/resources/blog/covidlock-update-coronavirus-ransomware
16. Schaab, P., Beckers, K., Pape, S.: A systematic gap analysis of social engineering defence mechanisms considering social psychology. In: 10th International Symposium on Human Aspects of Information Security & Assurance, HAISA 2016, Frankfurt, Germany, 19–21 July 2016, Proceedings (2016). https://www.cscan.org/openaccess/?paperid=301
17. Schaab, P., Beckers, K., Pape, S.: Social engineering defence mechanisms and counteracting training strategies. Inf. Comput. Secur. **25**(2), 206–222 (2017). https://doi.org/10.1108/ICS-04-2017-0022

A.10 Case Study: Checking a Serious Security-Awareness Game for its Legal Adequacy

Sebastian Pape and Dennis-Kenji Kipker. Case study: Checking a serious security-awareness game for its legal adequacy. unpublished manuscript, 2020

Case Study: Checking a Serious Security-Awareness Game for its Legal Adequacy

Sebastian Pape[1,2](✉)[0000−0002−0893−7856] and
Dennis-Kenji Kipker[3][0000−0003−1454−591X]

[1] Goethe-University, Frankfurt, Germany
[2] Social Engineering Academy (SEA) GmbH, Frankfurt, Germany
[3] University of Bremen, Bremen, Germany

Abstract. It is generally accepted that the management of a company has a legal obligation to maintain and operate IT security measures as part of the company's own compliance - this includes training employees with regard to social engineering attacks. On the other hand, the question arises whether and how the employee must tolerate associated measures, as for example social engineering penetration testing can be very intrusive. At a first glance, the decision to use a serious game for awareness raising and training against social engineering attacks, e.g. HATCH, seems to be fine. However, we investigate the legal challenges to make use of the game HATCH, which offers a possible application with a virtual and a real world scenario. As a case study, we examine under which circumstances which of HATCH's scenario types is suitable and legal to fulfill its goal. Based on the results, we derive general recommendations what to consider when making use of a serious game for awareness raising.

Keywords: Serious game · Labour law · Compliance · Social engineering · Awareness.

1 Introduction

Social engineering (SE) attempts to induce and exploit certain behaviour by influencing the victims to obtain sensitive information. A SE attack is often the first step of a larger attack, in which the attacker uses the information gained there for further attacks [2]. However, the latest Data Breach Investigations Report [2] also reports another increase of financially motivated SE, where the attacker directly ask for some money, i. e. by impersonating CEOs or other high-level executives. While a couple of defense methods and counteracting training methods [20, 21] exist, at present, companies have three main strategies to fend off SE attacks: SE penetration testing, security awareness training and campaigns.

For SE penetration testing, penetration testers are, as benign hackers, supposed to attack the employees and find weak points. This is mostly the case to

investigate the employees' vulnerability to phishing attacks. Unfortunately, this approach is not without problems. Experiments have shown that this approach can also lead to employees becoming demotivated when confronted with the results of the test [8]. In addition, such a test can interfere with the employees' right of personality, in particular since for an accurate assessment of the situation, employees cannot be told beforehand they are being tested, resulting in ethical issues [10]. As a consequence, there are numerous labour law requirements for SE penetration tests [14, 30].

Security awareness training may prove successful in particular against phishing. However, often employees are not trained at all or the training is conducted insufficiently [2] or in a way that it does not have a long lasting effect [25]. Security awareness campaigns often provide only information about risks and are not engaging, interesting and entertaining enough, evoke negative feelings such as anxiety, fear or stress and therefore are ineffective to change individuals' behavior [1]. Altogether, both strategies have in common that individuals generally dislike following instructions because it is associated with losing control.

A not so common method is the use of serious games, games that have a serious goal besides entertainment. Serious games are more entertaining and engaging than traditional forms of learning and influence individuals' behavior due to their use of pedagogy and game-based learning principles, such as motivation, cognitive apprenticeship and constructivism [9]. Therefore, at a first glance the use of a serious game for awareness raising and training against SE attacks, e. g. HATCH [4, 3], seems to be fine. However, in this paper we investigate the legal challenges to make use of the game HATCH, which offers two different types of scenarios. As a case study, we examine under which circumstances which of HATCH's scenario types is suitable and legal to fulfill its goal. Based on the results, we derive general recommendations what to consider when making use of a serious game for awareness raising.

2 Background and Related Work

In this section, we first describe HATCH, the game we have investigated. In the second part of the section, we discuss related work.

2.1 HATCH

The serious game considered for our use case is HATCH [4, 3], which aims to improve the employees' understanding of SE. For our analysis, we briefly sketch how HATCH works: Each player is in the role of an attacker.

1. Each player draws a card from the deck of *human behavioral principles*, e. g. the "Need and Greed" principle.
2. Each player draws three cards from the deck of the *social engineering attack techniques*, e. g. phishing.
3. Each player develops an attack targeting one of the personas in the scenario based on the drawn cards.

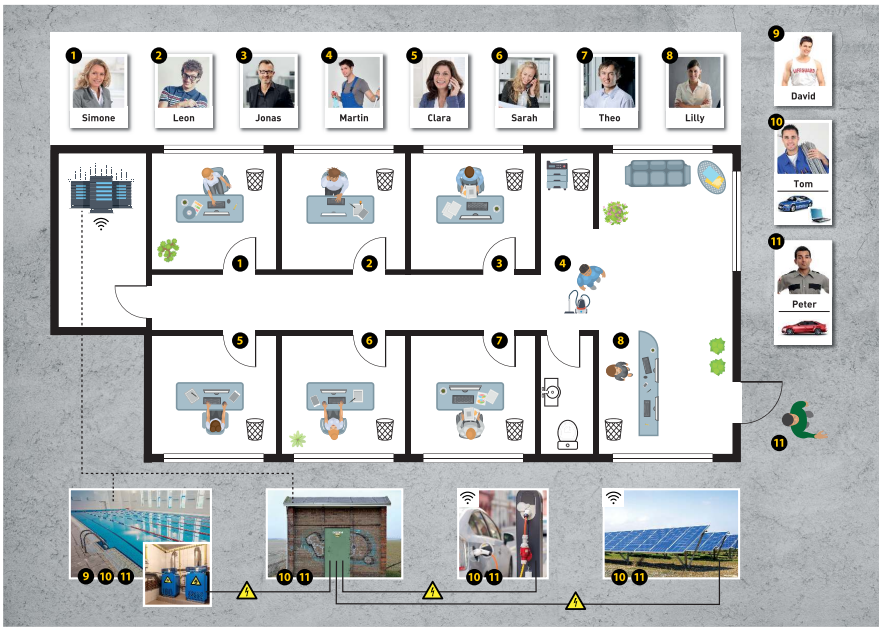


Fig. 1. Scenario for an Energy Provider

4. Each player presents his/her attack to the group and the other members of the group discuss if the attack is feasible.
5. The players get points based on how viable their attack is and if the attack was compliant to the drawn cards. The player with the most points wins the game.
6. As debriefing, the perceived threats are discussed and the players reflect their attacks.

The game can be played either with an imaginary (virtual) scenario or a (realistic) scenario that reflects the real working environment. We describe both scenario types in the following:

Virtual Scenarios Virtual scenarios are used when HATCH is used for training and awareness purposes [4]. These consist of a plan of a department or company (see Fig. 1) and for each of the employees shown in the plan there is a persona card that outlines the basic characteristics of the employee (see Fig. 2). The players' task now is to come up with an attack that is as plausible as possible on the basis of the drawn cards and that exploits the characteristics of the employees present in the game. The attack found is then evaluated for plausibility by the players.

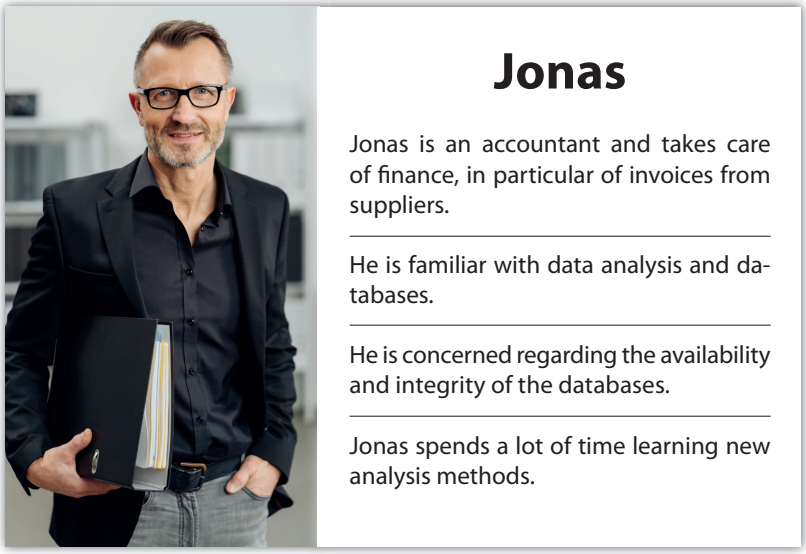


Fig. 2. Persona Card for Jonas, an Accountant

Realistic Scenarios The basic gameplay of HATCH with a realistic scenario [3] is the same as with a virtual scenario. However, virtual people are not used here, instead a plan of the real working environment is created and the players devise attacks on their colleagues. In doing so, they use their colleagues' existing knowledge of work processes, skills and preferences. Besides training and awareness raising, the result is a list of possible SE threats that can be used to improve work processes and security policies. The advantage over a threat analysis by experts is that the employees of a department or a company know the real work processes very well, so it is easier to train them in social engineering than to have experts study all work processes.

2.2 Related Work

While there are reports on the use of serious games in the corporate sector [9], the body of literature specific to serious games aiming to raise awareness and allow security training is rather low. Regarding compliance and serious games, there is a lot of work, but only on using serious games to increase the compliance and not on the compliance of serious games. In the area of SE, most of the work is focused on SE penetration testing. Hatfield [10] discusses the ethics of SE penetration testing, and Kuhn and Willemsen [14] and Zimmer and Helle [30] discuss SE penetration testing from a legal perspective towards labour law.

3 Legal Adequacy of HATCH

It is generally accepted that management has a legal obligation to maintain and operate IT security measures as part of the company's own compliance - this includes training employees with regard to social engineering attacks. The compliance obligation under IT security law can be derived from the most varied legal provisions and depending on the respective industry, generally from § 43 par. 1 German Limited Liablility Companies Act (GmbHG) and § 93 par. 1 German Stock Corporation Act (AktG). Where, on the one hand, there are corporate obligations to implement an appropriate level of IT security, the question arises on the other hand as to whether and how the employee must tolerate associated measures and, if necessary, also participate in them. The conflict between freedom and security is updated here in the form of issues relating to labour law and also data protection law, as well as for corporate compliance and corporate governance. Especially for an SE game like HATCH, which requires the active participation of the individual employee, various legal problem areas therefore open up. A distinction must be made between the realistic and the virtual game scenario.

3.1 Realistic Scenarios

In HATCH's realistic scenario, the actors involved in the company play themselves out. A particular legal relevance for this case arises from the fact that the simulated SE attacks are aimed at real persons and their character traits. With regard to the question of the legal reasonableness for the individual employee, this must be evaluated in compliance with Art. 2 par. 1 in conjunction with Art. 1 par. 1 of the German Constitution ("Grundgesetz", GG), which prescribes the General Right of Personality ("APR"). The APR as a part of the German Constitutional Law has an influence on employment law, among other things, as an ancillary obligation of the employer under the employment contract in accordance with § 241 par. 2 of the German Civil Code ("Bürgerliches Gesetzbuch", BGB).

For the employer, on the other hand, the freedom of occupation resulting from Art. 12 of the GG and the associated protection of entrepreneurial interests, also based on the indirect third-party effect of the fundamental rights in the private-law relationship, is in dispute. In principle, the employer must protect the employee from unlawful interference with his or her personal rights within the scope of his or her obligations arising indirectly from the APR [13, § 75 BetrVG, p. 99, Rn. 106]. This also includes protection against potentially embarrassing measures that could have a negative impact on employees [14, p. 112]. Particularly for an SE game in a realistic scenario, there are risks here in that employees feel exposed or that their company's appreciation is reduced, in that personal limits are exceeded by experiencing the game as a realistic situation and in that unforeseeable courses of the game occur in the group dynamics. It is questionable whether, in contrast to this and in the specific case, the company's interests in the execution of the game outweigh the risks and whether

compliance with the obligation under German IT Security Law is therefore to be classified as more important than employee protection. The principle applies here that in sectors and industries that are particularly relevant to security, gaps in corporate security certainly have a high weight in the legal weighing of interests [9]. From this, it can be concluded that, as a rule, the fictitious creation of a potentially employee-damaging environment, in which the real personality of the employee is exposed to weak points relevant to SE, in companies that are not particularly exposed, can hardly be justified by the potentially increased learning success of an awareness raising measure to promote IT security. The situation would be different for Critical Infrastructures with a high risk of attack or for companies that have already been victims of SE incidents and for which a similar threat situation is also apparent for the future: Here, the increased need for awareness-raising measures as a factual connection with the protection of employees and their jobs could justify the feasibility of the measure, above all in the interest of the employee. A different legal assessment may also be required in the case of a threat analysis, as the methodology to be applied here requires that all weak points relevant to IT security in a company be determined, which therefore necessarily also includes the human factor.

3.2 Virtual Scenarios

In the virtual scenario of HATCH, the SE attacks are played out using fictional characters and the imaginary role assignments associated with them. As in the realistic scenario, a legal balancing between the personal interests of the employee and the operational and economic interests of the employer must be carried out. A stigmatization risk for the individual employee exists here to the extent that technical or content-related knowledge gaps with regard to SE threats reveal personal deficits vis-à-vis the employer. However, this can be counteracted by training measures on SE prevention carried out before the game. Clearly formulated communication and game rules also help to ensure that situations of potential hostility, harassment or discrimination during the course of the game can be effectively countered in advance. Last but not least, the choice of fictional characters also significantly reduces the degree of personality impairment, as the employee's inner structures and characteristics are not subject to play [10]. Likewise, in the fictious scenario HATCH offers a possibility to promote and support the personality development of the employees within the scope of the compulsory exercise of § 75 par. 2 German Works Constitution Act ("Betriebsverfassungsgesetz", BetrVG). As in the realistic scenario, the game also enables the employer to protect the company from SE attacks by improving the awareness of its employees. As a result, the employer's interests generally outweigh those of the employee in the virtual game operation, so that the use of HATCH represents a conceivable alternative to the classic training measures in this area.

4 Discussion

In this section, we discuss how the result of our legal analysis could be generalised. First, which parts of the results can be transferred to other games. Second, to which extend it is possible to generalize the results to other (European) countries.

4.1 Generalization to Other Games

All legal considerations are specific to HATCH. Thus, in general one would need to do a legal assessment for each game individually. However, some general conclusion can be drawn in particular from the comparison of the two different scenario types. The analysis of the virtual scenario suggests that if within the serious game the employee's personal characteristics are not subject to play, the use of the serious game may be admissible if it is operated in a sufficient manner[4]. If the employee's personal characteristics are subject to play, as in the realistic scenario, a legal assessment is needed considering the aim of the game, i. ethreat analysis, the risk situation and exposure of the company to SE attacks to justify the feasibility of the game.

As a consequence, games which merely have a technical focus and do not consider human factors should be playable without the risk that employee's personal characteristics are subject to play. For example, *Elevation of Privileges* [22, 23] based on Shostack [24]'s threat modeling method should work out fine if players focus on the system, it's bugs and features as proposed in the game's instructions. Similar considerations hold for security related variants of planing poker [15] such as *Protection Poker* [27, 26], *Security Tactic Planning Poker (SToPPER)* [16].

Ctrl-Alt-Hack [5, 6, 7], another tabletop card game about white hat hacking, is based on game mechanics with virtual personas (hackers) and fulfilling the missions in the game does not rely on the players' or employees' characteristics. Therefore, even though it includes attacks based on social engineering, we would consider it comparable to the virtual scenario from HATCH, and thus conclude that there should be no major obstacles to play it within the context of a company.

We went through the descriptions of a couple of educational security games like *Cyber Security Requirements Education Game(SREG)* [29], *Cyber Security-Requirements Awareness Game (CSRAG)* [28], *Harbour Protection Table-Top Exercise (HPT2E)* [12, 11], *Operation Digital Chameleon* [19, 18], and *Operation Digital Snake* [17], but none of them was making use of players' or employees' characteristics. On the other hand, all of them are intended for awareness raising or education and none of them is intended for threat analysis. Thus, they would also be in the same line than the virtual scenario for hatch, which also makes them rather unproblematic game candidates.

[4] e. g. taking care that no personal deficits vis-á-vis the employer are revealed and clearly formulated communication and game rules are applied

4.2 Generalization to Other Countries

All legal considerations made in this context are subject to German law. This is due to the fact that in the EU, labour law is primarily regulated by the Member States themselves. Nevertheless, some general conclusions can also be drawn. For example, some of the legal considerations made in the legal analysis in this article are based on data protection regulations which are governed by EU law, in particular the EU GDPR. In many cases of EU law, as far as the processing of personal data is concerned, the focus is on balancing the interests of the data processor (in this case, the employer) and those whose personal data are processed (in this case, the employee). Thus, to the extent that operational IT security interests are weighed against individual data protection interests, the legal statements in this paper can certainly be generalised to a certain extent. In this respect, the legal weighing of interests carried out here can at least provide an indication of whether the use of HATCH in the operational context would also be legally permissible in other (European) countries.

5 Conclusion

While at a first glance, it seems to be legit to use a serious game for security training and awareness, our legal assessment showed large differences in the assessment of the two different scenario types. If the employee's personal characteristics are part of the game, care needs to be taken to not unnecessarily expose the personality of the employees. This even holds if the employees ask for or volunteer to play the scenario with a realistic environment, where they would suggest social engineering attacks on each other. On the other hand, if the employer can demonstrate a reasonable interest, i .e. if the game is used for threat analysis, the use of the game with a realistic scenario may be admissible.

As future work, the legal assessment should be extended for other countries such as the US or other member states of the EU.

Acknowledgement

This work was supported by European Union's Horizon 2020 research and innovation program from the project CyberSec4Europe (grant agreement number: 830929) and from the project THREAT-ARREST (grant agreement number: 786890). We are thankful to Kristina Femmer for the graphical implementation of the scenario and the persona cards.

Bibliography

[1] Bada, M., Sasse, A.M., Nurse, J.R.C.: Cyber security awareness campaigns: Why do they fail to change behaviour? CoRR abs/1901.02672 (2019), http://arxiv.org/abs/1901.02672

[2] Bassett, G., Hylender, C.D., Langlois, P., Pinto, A., Widup, S.: Data breach investigations report (2020), https://enterprise.verizon.com/resources/reports/2020-data-breach-investigations-report.pdf

[3] Beckers, K., Pape, S.: A serious game for eliciting social engineering security requirements. In: Proceedings of the 24th IEEE International Conference on Requirements Engineering. RE '16, IEEE Computer Society (2016)

[4] Beckers, K., Pape, S., Fries, V.: HATCH: Hack and trick capricious humans – a serious game on social engineering. In: Proceedings of the 2016 British HCI Conference, Bournemouth, United Kingdom, July 11-15, 2016 (2016), https://ewic.bcs.org/content/ConWebDoc/56973

[5] Denning, T., Kohno, T., Shostack, A.: Control-alt-hack^{TM}: a card game for computer security outreach and education (abstract only). In: The 44th ACM Technical Symposium on Computer Science Education, SIGCSE '13, Denver, CO, USA, March 6-9, 2013. p. 729 (2013), http://doi.acm.org/10.1145/2445196.2445408

[6] Denning, T., Lerner, A., Shostack, A., Kohno, T.: Control-alt-hack: the design and evaluation of a card game for computer security awareness and education. In: 2013 ACM SIGSAC Conference on Computer and Communications Security, CCS'13, Berlin, Germany, November 4-8, 2013. pp. 915–928 (2013), http://doi.acm.org/10.1145/2508859.2516753

[7] Denning, T., Shostack, A., Kohno, T.: Practical lessons from creating the control-alt-hack card game and research challenges for games in education and research. In: 2014 USENIX Summit on Gaming, Games, and Gamification in Security Education, 3GSE '14, San Diego, CA, USA, August 18, 2014. (2014), https://www.usenix.org/conference/3gse14/summit-program/presentation/denning

[8] Dimkov, T., Van Cleeff, A., Pieters, W., Hartel, P.: Two methodologies for physical penetration testing using social engineering. In: Proceedings of the 26th annual computer security applications conference. pp. 399–408 (2010)

[9] Donovan, L., Lead, P.: The use of serious games in the corporate sector. A State of the Art Report. Learnovate Centre (December 2012) (2012)

[10] Hatfield, J.M.: Virtuous human hacking: The ethics of social engineering in penetration-testing. computers & security 83, 354–366 (2019)

[11] Kessel, R., Gwatkin, N.: Harbour protection table - top exercise hpt2e 20 - 23 march 2012, la spezia : Hpt2e technologies and platforms. https://www.cmre.nato.int/research/publications/all-publications/technical-reports/special-publications/447-harbour-protection-table-top-exercise-hpt2e-20-23-march-2012-la-spezia-hpt2e-technologies-and-platforms (April 2012), nURC special report ; SP 2012 002, FIZBw-Bestellnummer DS 2262, Reportnummer NURC SP 2012 002

[12] Kessel, R., Gwatkin, N.: Harbour protection table-top exercise hpt2e: Contextual read ahead. https://www.cmre.nato.int/research/publications/all-publications/technical-reports/special-publications/446-harbour-protection-table-top-exercise-hpt2e-

contextual-read-ahead (April 2012), nURC special report ; SP 2012 001, FIZBw-Bestellnummer DS 2263, Reportnummer NURC SP 2012 001
[13] Kreutz: GK-BetrVG, Bd. 2. 10 edn. (2014)
[14] Kuhn, J., Willemsen, A.: Arbeitsrechtliche Aspekte von Social Engineering Audits. DER BETRIEB 02, 111–117 (2016), https://www.wiso-net.de/document/MCDB__DBDBDB1167400
[15] Moløkken-Østvold, K., Haugen, N.C., Benestad, H.C.: Using planning poker for combining expert estimates in software projects. Journal of Systems and Software 81(12), 2106–2117 (2008)
[16] Osses, F., Márquez, G., Orellana, C., Astudillo, H.: Towards the selection of security tactics based on non-functional requirements: Security tactic planning poker. In: 2017 36th International Conference of the Chilean Computer Science Society (SCCC). pp. 1–8. IEEE (2017)
[17] Rieb, A.: Kma homepage article about operation digital snake game. https://www.kma-online.de/aktuelles/it-digital-health/detail/klinik-mitarbeiter-wappnen-sich-mit-planspielen-gegen-cyberattacken-a-38128/2 (2018)
[18] Rieb, A., Lechner, U.: Operation Digital Chameleon – Towards an Open Cybersecurity Method. In: Proceedings of the 12th International Symposium on Open Collaboration (OpenSym 2016). pp. 1–10. Berlin (2016), http://www.opensym.org/os2016/proceedings-files/p200-rieb.pdf
[19] Rieb, A., Lechner, U.: Towards Operation Digital Chameleon. In: Havârneanu, G., Setola, R., Nassopoulos, H., Wolthusen, S. (eds.) CRITIS 2016 - The 11th International Conference on Critical Information Infrastructures Security (to appear). pp. 1–6. Paris (2016)
[20] Schaab, P., Beckers, K., Pape, S.: A systematic gap analysis of social engineering defence mechanisms considering social psychology. In: 10th International Symposium on Human Aspects of Information Security & Assurance, HAISA 2016, Frankfurt, Germany, July 19-21, 2016, Proceedings. (2016), https://www.cscan.org/openaccess/?paperid=301
[21] Schaab, P., Beckers, K., Pape, S.: Social engineering defence mechanisms and counteracting training strategies. Information and Computer Security 25(2), 206–222 (2017), https://doi.org/10.1108/ICS-04-2017-0022
[22] Shostack, A.: Elevation of privilege: Drawing developers into threat modeling. Tech. rep., Microsoft, Redmond, U.S. (2012), http://download.microsoft.com/download/F/A/E/FAE1434F-6D22-4581-9804-8B60C04354E4/EoP_Whitepaper.pdf
[23] Shostack, A.: Elevation of privilege: Drawing developers into threat modeling. In: 2014 USENIX Summit on Gaming, Games, and Gamification in Security Education (3GSE 14). USENIX Association, San Diego, CA (Aug 2014), https://www.usenix.org/conference/3gse14/summit-program/presentation/shostack
[24] Shostack, A.: Threat Modeling: Designing for Security. John Wiley & Sons Inc., 1st edn. (2014)
[25] Stahl, S.: Beyond information security awareness training: It's time to change the culture. Information Security Management Handbook, Volume 3 3, 285 (2006)

[26] Williams, L., Meneely, A., Shipley, G.: Protection poker: The new software security "game". Security Privacy, IEEE 8(3), 14–20 (May 2010)
[27] Williams, L., Gegick, M., Meneely, A.: Protection poker: Structuring software security risk assessment and knowledge transfer. In: Proceedings of International Symposium on Engineering Secure Software and Systems. pp. 122–134. Springer (2009)
[28] Yasin, A., Liu, L., Li, T., Fatima, R., Jianmin, W.: Improving software security awareness using a serious game. IET Software (2018)
[29] Yasin, A., Liu, L., Li, T., Wang, J., Zowghi, D.: Design and preliminary evaluation of a cyber security requirements education game (sreg). Information and Software Technology pp. – (2017), https://www.sciencedirect.com/science/article/pii/S0950584917301921
[30] Zimmer, M., Helle, A.: Tests mit Tücke - Arbeitsrechtliche Anforderungen an Social Engineering Tests. Betriebs-Berater 21/2016, 1269 (2016)
All urls have been last visited on February 12th, 2021.

Appendix B

Security Management

B.1 Defining the Cloud Battlefield – Supporting Security Assessments by Cloud Customers

Sören Bleikertz, Toni Mastelic, Sebastian Pape, Wolter Pieters, and Trajce Dimkov. Defining the cloud battlefield – supporting security assessments by cloud customers. In *Proceedings of IEEE International Conference on Cloud Engineering (IC2E)*, pages 78–87, 2013. doi: 10.1109/IC2E.2013.31

Defining the Cloud Battlefield
Supporting Security Assessments by Cloud Customers

Sören Bleikertz*
IBM Research - Zurich
sbl@zurich.ibm.com

Toni Mastelić*
Vienna University of Technology
toni@infosys.tuwien.ac.at

Sebastian Pape*
TU Dortmund
sebastian.pape@cs.tu-dortmund.de

Wolter Pieters
TU Delft / University of Twente
w.pieters@tudelft.nl

Trajce Dimkov
Deloitte LLP
tdimkov@deloitte.nl

***Abstract*—Cloud computing is becoming more and more popular, but security concerns overshadow its technical and economic benefits. In particular, insider attacks and malicious insiders are considered as one of the major threats and risks in cloud computing. As physical boundaries disappear and a variety of parties are involved in cloud services, it is becoming harder to define a security perimeter that divides insiders from outsiders, therefore making security assessments by cloud customers more difficult.**

In this paper, we propose a model that combines a comprehensive system model of infrastructure clouds with a security model that captures security requirements of cloud customers as well as characteristics of attackers. This combination provides a powerful tool for systematically analyzing attacks in cloud environments, supporting cloud customers in their security assessment by providing a better understanding of existing attacks and threats. Furthermore, we use the model to construct "what-if" scenarios that could possible lead to new attacks and to raise concerns about unknown threats among cloud customers.

I. Introduction

A. Background: Security Concerns in Cloud Computing

Cloud computing has gained remarkable popularity in recent years due to the economic and technical advantages of this new way of delivering computing resources. Customers benefit from rapid provisioning and seemingly infinite scalability, while only being charged on a pay-per-use basis.

Although the benefits of cloud computing are evident and users demand cloud services, security is a major inhibitor [1]. An analysis of risks and threats in cloud computing has been conducted in [2] and [3]. In particular, both reports agree that insider attacks and malicious insiders are a major technical risk and among the top 10 threats. The risk is amplified due to the disappearance of physical boundaries that makes it very challenging to define a security perimeter that divides insiders from outsiders [4], [5].

Due to the variety of parties involved in a cloud service, cloud customers face difficulties in assessing the risks and threats of insider attacks in cloud services. To illustrate this point, let us consider the following attack scenarios: A malicious cloud administrator can steal information that are stored or processed in a virtual machine of a cloud customer [6]. Furthermore, a malicious cloud customer can perform a similar attack on other customers that share the same physical resources [7]. Malicious behavior is not always required to constitute a risk to cloud customers. The outage of Amazon EC2 in 2011 [8] impacted the availability of the cloud service and was caused by an honest fault of a cloud administrator. Similar, honest faults by cloud customers can also impact other customers as demonstrated in [9], where a SSH public key for the administrator account was accidentally left in an image and which constituted a backdoor.

These scenarios cover only a small set of the involved parties (i.e., only cloud administrator and customer) and just two different characteristics of the attacker (i.e., honest faults and malicious). However it shows that the general misconception of either trusting the cloud or not does not hold, but more fine-grained trust and attacker models are required. We need to systematically specify the parties, their capabilities and motivations, in order to obtain a complete picture and support cloud customers in their risk and threat assessments.

B. Research goal: Supporting Security Assessment of Infrastructure Clouds

In this paper we propose a high-level model that supports cloud customers in their security assessments of clouds. Since the security of a cloud strongly depends on the used infrastructure, our framework combines a system model of infrastructure clouds, including the involved entities and system components, with a security model that describes security objectives of cloud customers, attacker characteristics, and threats. The framework allows for systematic analysis of the security threats in a specific cloud service environment. By comparing the analysis across cloud providers, decisions on provider choice are supported.

The main challenges involved are related to reaching the appropriate level of abstraction. In theory, many different entities could be distinguished in the model, but this comes at the cost of increased complexity. As the model is meant to be used by cloud customers, understandability and usability

* These authors have contributed equally and are ordered alphabetically by their surnames.

are important requirements. In addition, it turns out that several unique features are essential in modeling existing attack scenarios, for instance access intervals, i.e., when an entity can access certain resources. Finding a combination of expressiveness and understandability is therefore key.

C. Methodology: Designing an IaaS Threat Model

Our model focuses on infrastructure clouds (IaaS) as the most generic and standardized abstraction layer [10]. In many cases the layers build upon each other, therefore a model of IaaS also partly covers attack scenarios of Platform-as-a-Service (PaaS) and Software-as-a-Service (SaaS).

We develop our system model starting from entities, system components, and their relation in terms of access levels for infrastructure clouds. We consider further entities besides the cloud provider and the customer, such as hardware manufacturers. Thereby our model is able to cover an extended set of possible attacks, for example also hardware trojans [11].

As the model is meant to support security assessment by cloud customers, the security objectives in our security model are defined from a customer's point of view. For now, we focus on confidentiality, integrity, and availability *(CIA)* with regard to compute, storage, and network resources provided by an infrastructure cloud provider. Unlike previous work, we differentiate between different characteristics and motivations of attackers in our security model, ranging from a *malicious* attacker to a *stepping stone* one, who unknowingly contributes to an attack. This allows us to assess whether implemented measures match the expected type of attackers. The combination of (a) our system model, (b) the security objectives of cloud customers, and (c) our attacker model forms the basis of our threat model, which can be used to analyze and identify attack scenarios.

For the evaluation of our model, we mapped existing practical attacks in cloud environments to the model by identifying the involved entities, their attacker characteristics, and threats. We performed several iterations of model development and evaluation: After each iteration we improved the model based on our findings when mapping the attacks.

For a systematic analysis of threats in cloud environments we propose a variation strategy inspired by the HAZOP approach [12]: First we form the foundation of our analysis by identifying known attacks and mapping them to the model. Second, we analyze remaining combinations of entities, attackers' characteristics, and threats in order to reveal possible unknown attacks. Due to space restrictions we only give derivations of a subset of possible new attack scenarios in this paper.

D. Our Contributions

In summary, we make the following contributions:

- We propose a comprehensive system model of infrastructure clouds.
- Our security model defines security objectives, as well as a set of archetypes that can capture a wide range of characteristics and motivations of an attacker.
- The combination of our security model and system model provides a powerful tool for systematically analyzing existing attacks in cloud environment. We demonstrate this by a set of known attack scenarios.
- Finally, our model can be used for deriving new security threats from existing scenarios, as well as describing and analyzing new what-if scenarios by changing characteristics of involved parties.

II. Related Work

In comparison to existing work by Abbadi et al. [43], our model is more comprehensive for infrastructure clouds due to our focus on this abstraction layer, while their model is more abstract and covers also other layers, i.e., SaaS. In [44] Grobauer and Walloschek focus on risk assessment and vulnerabilities of technologies used in a cloud environment. They correlate these vulnerabilities to essential cloud characteristics and to system components like computational resources and storage. However, they do not consider involved parties and their relations to system components, nor do they try to provide a model for mapping these vulnerabilities to attack scenarios. Similarly, Garfinkel and Rosenblum [45] discuss security problems at the virtualization layer that now forms an integral part in infrastructure clouds. However, their work predates the cloud computing paradigm and does not discuss such security issues in a larger scenario that also considers the variety of entities and possible attackers found in infrastructure clouds. Behl [46] addresses the most common and critical security issues in cloud computing and provides key research challenges in this field. Although his work covers insider and outsider attack scenarios, they are discussed as separate use-cases, while no model is provided to correlate and describe them. A survey of threats in cloud environment is presented by Molnar and Schechter [47], although they do not claim to provide a necessarily complete set of threats and the authors expect that new threats will be identified. We believe that our model can be used to identify and contribute new threats due to our systematic approach, as well as to provide a categorization of existing ones.

III. System Model

Cloud computing can be implemented on different abstraction layers ranging from the lowest level of Infrastructure-as-a-Service *(IaaS)* to the highest abstraction of Software-as-a-Service *(SaaS)*. Developing a generic threat model covering all the abstraction layers is hard, since we have to deal with an increasing diversification on the higher abstraction layers. For example, both Google GMail and Salesforce CRM are considered instances of SaaS, but with different and application-specific attacker models. Therefore, we define a model of a cloud environment on a IaaS layer consisting of entities and the system components as shown on Figure 1.

A. Entities

Entities represent subjects which are involved in a cloud service, directly or indirectly, while components represent

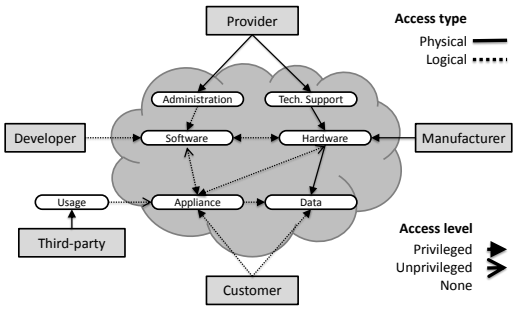


Figure 1. System model with relations between entities and components.

objects of which a cloud service is composed of. Entities include:

Provider - entity providing a cloud service by managing and operating a cloud infrastructure, which includes hardware and software resources.

Manufacturer - entity producing a hardware resource that is being used by the *provider* as part of the cloud service. The *provider* chooses a manufacturer from which it will acquire hardware for its cloud service.

Developer - entity producing a software resource that is being used by the *provider* as part of the cloud service. The *provider* chooses a *developer* from which it will acquire software for its cloud service.

Customer - user of the cloud service provided by the *provider* that uses *software* and *hardware* resources as part of that service. The *customer* chooses a *provider* whose services he will be using.

Third-party - entity not directly involved in providing or using an IaaS service, but can represent user on higher layers of the cloud service (e.g., SaaS). The *third-party* can choose an IaaS *customer* whose upper layer service he will be using.

B. Components

Each entity has one or more components, which can be accessed physically or logically. All components and their access types are shown in Figure 1 and are explained below:

Administration - a management and operational service provided by a *provider* with logical access to the software infrastructure.

Technical Support - a management and operational service provided by a *provider* with physical access to the hardware infrastructure.

Hardware - products like hard-disk, processor, network switch etc. produced by a *manufacturer*, and used as part of a cloud data center.

Software - products like hypervisor, cloud management software etc. produced by a *developer*, and used as part of a cloud infrastructure.

Data - information stored on a hardware or being transmitted.

Appliance - an executable piece of software deployed by a *customer* using a cloud service. It represents the higher layer of a cloud service, e.g., SaaS, thus it is considered as a black box completely controlled by a *customer*. It is managed by cloud management software, while it can be logically accessed by a *third-party*. Appliances that are not running are considered as *data*, e.g., a virtual machine image stored on a disk.

Usage - component representing the usage by a *third-party* entity, which is not directly involved in the cloud service. It can logically access an appliance deployed by a *customer*.

Upper layers of a cloud service are covered with the *appliance* component since it is under full responsibility of a *customer*. A *customer* chooses either a preconfigured *appliance* from another *customer*, or chooses software components (operating system, applications, etc.) and assembles/configures the *appliance* according to his needs. In both cases, the *customer* needs to assess the third-party software – either in the form of a preconfigured appliance, or as individual software components. Thereby, we treat the *appliance* as a blackbox, otherwise the cloud model would be stretched beyond the targeted scope (IaaS).

However, each entity or component can have multiple instances when used for describing an attack scenario, e.g., there can be several *customers*, each of them having their own *appliances*; or a *provider* can buy pieces of hardware from different manufacturers, thus having several instances of a *manufacturer* entity, as well as several instances of a *hardware* component.

C. Access Level

The relationship between entities and their components, as well as between components themselves, is defined through an access level. An access level represents a level of privileges one entity or component has over another. Figure 1 shows access

levels between components represented by different types of arrows:

privileged - full access with all the privileges for configuring and manipulating a component.
unprivileged - limited access to functionality or an interface of a component.
none - no access.

Access level has two attributes: direction and transitivity. If A has a privileged logical access to B, it doesn't imply that B has the same type and level of access to A, which is defined by the direction attribute, e.g., *hardware* component having a privileged physical access to *data*, while *data* has no access to *hardware* as it is simply stored on it. Transitivity defines that A can use its access to B in order to manipulate C, where B has access to C. For example, *administration* can use *software* to manipulate *appliance*. Additionally, a certain access level can be changed by obtaining more privileges, e.g., an attacker can use his unprivileged access level to exploit a vulnerability in a component and acquire privileged access to that component.

According to the above classification, the access level between entities and their components is always considered as *privileged* since the entity owns the component. However, more fine-grained access levels between entities and components depend on how often can an entity access its component:

One-time - an entity can access a component only once, i.e., a *manufacturer* can physically access *hardware* only when it is being produced.
Periodic - an entity can access a component on periodic bases, i.e., a *developer* can logically access *software* (i.e., hypervisor) only when the *software* is being updated after it has been deployed. Note that the idea of periodic access levels is not that the entity necessarily has access at a certain point in time (e.g. each Monday), but rather a recurrent and non-continuous access.
Permanent - an entity can access a component at any moment and all the time, i.e., a *provider* can typically perform *administration* at any time.

Our definition of access levels implicitly forms a hierarchy of entities based on access privileges and their attributes (cf. "Insiderness" [13]).

IV. Security Model

In this section we define the security objectives for cloud customers and the attacker model with its different attacker characteristics. Moreover, we define our threat model that combines the system model, security objectives, and attackers.

A. Security Objectives of Cloud Customers

The security objectives in our security model are defined from a cloud customer's point of view. Our primary concern is the exposure of sensitive business or personal information belonging to the customers of a cloud provider. For now, we are focusing on confidentiality, integrity, and availability *(CIA)* with regard to computing, storage, and network resources provided by an infrastructure cloud provider. We define the following security objectives with regard to the components defined in Section III.

Confidentiality of:

- *S1* Appliances when executed.
- *S2* Data when stored.
- *S3* Data and appliances when transmitted over a network.

Integrity of:

- *S4* Appliances when executed.
- *S5* Data when stored.
- *S6* Data and appliances when transmitted over a network.
- *S7* Software: Hypervisor and management software remain in a "good" state (e.g., no backdoors will be installed).
- *S8* Hardware: Remains in a "good" state.

Availability of:

- *S9* Appliances: both for owning customers and third parties, who consume services provided by appliances.
- *S10* Data: both for customers and appliances accessing data.
- *S11* Software: management infrastructure and hypervisor remain functional.
- *S12* Hardware (analog to Software).

The security objectives *S7*, *S8*, *S11*, and *S12* are correlated to others, i.e., once they are not achieved, it is likely that the other cloud customer specific objectives will also not be achieved. Note that other common security objectives such as the theft of computational resources are covered by *S4* as in many cases the integrity of the appliance has to be violated (e.g. by installing malware) before the appliance can be abused for the attacker's purposes.

B. Attacker Model

Parties participating in cloud services may be characterized along two dimensions: goals and skills. Goals specify what a party wants to achieve and skills specify the ability of a party to realize these goals.

To specify *goals*, utility functions are typically employed from an economic point of view [14], [15]. Such functions map the outcomes of attack scenarios to a single-scale (typically monetary) value for the party involved. Different inputs can contribute to the utility, such as damage caused (for terrorist attackers), expected gain, costs, and risks associated with the scenario [15]. Utility functions do not only apply to attackers, but also to honest entities. For example, a *cloud provider* that cares about its *customers* will have negative utility associated with damage to *customers*.

Skills describe the abilities of parties to realize these goals, and typically determine the outcome of scenarios when different parties have conflicting goals. For example, when a *cloud provider* aims to secure its systems against disruption, but has low skill, and a terrorist attacker aims at disruption the service, with high skill, the likelihood of disruption will be determined by the difference in skill levels [16]. Skill level can be further

divided to include a notion of available resources, but we will not use that here.

Archetypes combine goals and skills. Different archetypes of contributors to an attack scenario may be defined:

malicious (intentionally contribute to an attack): the entity intends to increase risk and associated damage to other entities for its own gain;

ostrich (knowingly contribute to an attack): the entity does not intend to increase risk for others, but fails to take action upon being informed about this;

charlatan (failing to acquire essential knowledge about contributing to an attack): the entity increases risk for others, does not know about this, but could/should have known;

stepping stone (unknowingly contribute to an attack): the entity actually increases risk for others, but could not have known.

The malicious and ostrich archetypes are driven by goals, e.g., causing damages or for monetary reasons, and their skill level determines the success of reaching such goals. The charlatan and stepping stone archetypes have low skills, which renders their goal of providing a secure cloud service to their customers unsuccessful. The ostrich can also been called lazy, and the term sloppy can been used for charlatans and stepping stones. Moreover, there may be an additional archetype involved, which does not have the characteristic of an attacker:

defender (actively tries to prevent an attack): the entity reduces the risk for others, e.g. by increasing the burden of a successful attack. The motivation for a defender may for example be that he is a reputationalist (who tries to improve utility of others to maintain reputation and thereby its own utility) or an altruist (who tries to improve the utility of others without necessarily benefiting itself; cf. corporate social responsibility).

Defined archetypes are applied on entities, while components inherit the archetypes from them. Archetype inherited from an entity represents a best possible archetype a component can have, while it still can have a worse one, e.g., *provider* can be a *charlatan*, which means that an *administration* can only be *charlatan* or worse, i.e., *ostrich* or *malicious*.

C. Threat Model

In order to describe or assess a certain threat, we must include all entities and components involved in the attack. Moreover, each entity is characterized with an archetype, a combination referred to as a *role*, e.g., *ostrich provider* or a *malicious usage*. Along with involved components, a role represents a building block of a *scenario* where roles are often combined, e.g., a *charlatan provider* plus a *malicious technical support*. A scenario thus describes how entities with certain archetypes behave towards the system in a specific setting, thereby setting the scene for an attack. For example, the above scenario with a charlatan provider and a malicious technical support may result in certain data being leaked.

After defining a scenario by using a system model defined in Section III and archetypes from Section IV-B, we combine it with a security objective (Section IV-A) in order to analyze a *threat* as shown in Figure 2. A threat thus signals that a particular scenario may violate a particular security objective through an attack. For example, the above data leaking scenario constitutes a confidentiality threat.

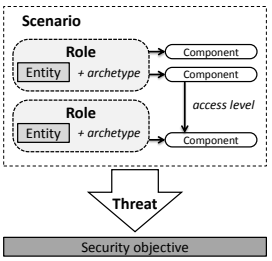


Figure 2. Deriving a threat from a role based scenario and security objective.

The likelihood of a threat is influenced by an attacker entity's access levels, including the access interval, as well as the characteristics (including skills and goals) of the attacker.

V. Model Applications

To check the usability and generality of the model, we assembled a set of security threats from the Cloud Security Alliance [17], ENISA [18], and the Deloitte Cloud Risk Map [19]. For each of the threats, we developed attack scenarios using subsets from the proposed model. This exercise helped in iteratively improving the constructs in the model. Some of the lessons learnt during the development of the scenarios are presented in the conclusions part of the paper.

The model can be used for several practical purposes. First, the model can explain the success of existing attacks, and possible mitigations. Second, the model can produce a systematic set of threats by examining each of the entities and each relation between the entities. Having such an extensive list of threats is an important input in developing a security assessment for a cloud solution. Third, the model can be used to analyze the behavior of all the entities participating in the cloud solution and their possible motivation behind their behavior. Such analysis would provide insights into the causes of threats in addition to a cost-benefit assessment. Finally, the model can be used to define possible attack scenarios by presenting what-if scenarios in a consistent language. What-if scenarios are useful in penetration tests on cloud solutions, as they expose possible design vulnerabilities in the solution.

A. Applying the Model to Practical Attacks

This section provides detailed sample scenarios which are used to illustrate the definitions before and show how our model can be applied to already well known attack scenarios.

The described attacks are an evaluation of our model and demonstrate that the model is sufficient to cover different scenarios described in the literature or which already exist in

the real world. On the other hand, the application of the model can be used to identify threats or derive new possible attack scenarios.

1) Malicious Administrator Attacks:

a) Scenario Description: Cloud computing is fundamentally based on the virtualization of servers. This means that the administrators managing the servers should carefully be selected, since they are powerful insiders. There exist several known attacks which the administrators of such servers could try to mount. Oberheide et. al. [20] propose an attack on VMWare or Xen that targets the live migration of virtual machines where a virtual machine is transferred to another host without halting it. As a proof of concept for man-in-the-middle attacks during the migration of a virtual machine, they show the possibilities of changing memory data or injecting an SSH authentication key during migration. With a similar idea, Rocha and Correia [6] demonstrate attacks by an administrator with root access on the hypervisor, but no access on the virtual machine itself. By making use of memory dumps or images of the (virtual) hard drive they show how to derive clear text passwords or cryptographic keys.

b) Model Application: The basic principle of the attack is shown in Fig. 3. The *malicious provider* or a *malicious administrator* accesses the attacked *appliance* via his privileged access on the *software* layer. Note that the provider itself may be *malicious* or he may be in the range from *ostrich* to *stepping stone* and thus hired untrustworthy administrators resulting in an *malicious administration*. Regarding the memory dumps and corruptions, the *appliance's* memory can be read or written during *administration* and thus the confidentiality and integrity of the running *appliance* is violated (*S1*, *S4*). The *administration* is also able to read or corrupt the *appliance's* template when it is stored (*S2*, *S5*) or transmitted over the network (*S3*, *S6*). The remaining security objectives regarding the hypervisor are affected on the *software* layer (*S7*), but not on the *hardware* layer (*S8*), since only *technical support* has access to the hardware. All *administration* tasks are in general granted privileged access, and they may shutdown the *appliance* or the underlying *hardware*, therefore violating all security objectives regarding availability (*S9* - *S12*). At first glance, it seems that the *administration* has permanent access, but the *administration* may have only periodic access, since the tasks may follow a certain schedule and extra cycles might raise suspicion.

c) Mitigation and Assessment: Although the functional difference between the possible archetypes of the provider are not apparent, because they all heighten the risk of vulnerability for the cloud customer, from an overall risk management perspective they make a difference. When the customer evaluates mitigation strategies for their overall security assessment, different methods and processes protect against the different archetypes. For example, a *charlatan provider* hires a *malicious administrator*, because the necessary background checks are not implemented in the hiring process of the provider. A cloud customer can verify the existence of such processes during their security assessment of a cloud provider. Similarly,

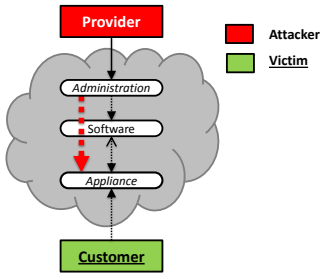


Figure 3. Malicious administration manipulating an appliance.

a *charlatan provider* fails to implement proper handling of security vulnerability reporting, or an *ostrich* one does not perform necessary patch management once being informed about a vulnerability. Besides processes, there also exists technical mitigation possibilities. Trusted hypervisors [21], [22] or access control approaches [23] can protect against *malicious* administrators. Fully homomorphic encryption [24] enables computations on encrypted data, but it is still practically infeasible [25]. A two-person administration [26] may mitigate faults by *charlatan administration*.

2) App Store Scenario:

a) Scenario Description: In a cloud app store scenario *customers* (publishers) offer *appliance* templates containing software applications to the other *customers*. Referring to Wei et. al. [27] there are two main risks in an app store scenario. On the one hand, the publisher may have inserted malware such as a Trojan horse in the provided appliance. Since it is crucial for the *provider* to maintain the reputation of his cloud app store, the *provider* tries to prevent the distribution of malware, for example by scanning the provided appliances.

On the other hand, the publisher may reveal sensitive information, especially when releasing pre-configured appliances. Bugiel et. al. [9] describe how they were able to automatically extract sensitive information– such as Amazon Web Service API keys, private keys and login credentials, private data and source code. Again, the cloud *provider* tries to prevent this by giving warnings in its user guide [28] or by disabling affected appliances.

b) Model Application: The relevant entities for modeling the two attacks described before are the *provider* and two different instances of *customers*. The publisher and the user of the provided *appliance* are both instances of the entity *customer*. While the *provider* is only watching or guarding its *customers*, the two *customers* attack each other at the appliance level (cf. Fig. 4) by either providing *appliances* with malware or finding sensitive information in the provided *appliances*. Therefore one of the *customers* is a malicious attacker and the other *customer* is the victim of the attack.

Regarding the distribution of *appliances*, the concerned security objective is mainly the leak of confidential information

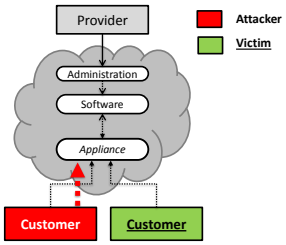


Figure 4. Attacking other customers through appliances.

(*S2*) as a direct consequence. However, depending on the leaked information as an indirect consequence, e.g., leakage of login credentials, the attacker may be able to access the victim's *appliance* and thus additionally threatens the *customer*'s availability (*S9*, *S10*). Moreover, the integrity of computations and stored data is threatened (*S4*, *S5*) and the confidentiality of the *customer*'s computations (*S1*). In the case of malware in *appliances*, the attacker may directly gain access to the *appliances*, and thus all mentioned security objectives (*S1*, *S2*, *S4*, *S5*, *S9*, *S10*) are violated.

Since the *provider* is making the *appliances* available via its app store, the characteristics of the *provider* may be *ostrich* (if the *provider* knows there are *appliances* with malware or sensitive information in its app store), *charlatan* (if the *provider* simply does not care which *appliances* are provided in its app store but perhaps has a marketing team promising excellent quality of the provided *appliances*) or *stepping stone* (if the *provider* just does not know about the problems within the *appliances*).

c) Mitigation and Assessment: In the example given in the scenario above, Amazon first was a *stepping stone*, since they stated that they do not check the *appliances*, but then changed their characteristics to *defender* (reputationalist) since they informed affected *customers* and removed concerned *appliances* from their app store. This approach represents post-emptive measure, which requires scanning and cleaning of infected/malicious images [29]. However, instead of cleaning the VM image repository, a *provider* can implement a pre-emptive image management system that provides a secured access to images [27]. Additionally, a *defender provider* could also perform patching of VM images in order to provide up-to-date security measures for his images [30].

3) Side-channel Attacks:

a) Scenario Description: The setup of a side-channel attack scenario consists of a *customer* who tries to attack another *customer* by placing a virtual machine on the same physical server and trying to observe the system's behavior. Ristenpart et. al. [7] demonstrated such an attack on the Amazon EC2 infrastructure. They show how to map the internal cloud structure, and identify where the virtual machine of the victim is likely to reside. The attacker may then instantiate a virtual machine which is located on the same physical machine as the virtual machine of his victim. Using such a co-located virtual machine, the attacker then try to mount side-channel attacks across the boundaries of the virtual machines. Ristenpart et al. referred to cache-based side channels. For example, they demonstrated how to estimate the load of the underlying physical machine, which indicates activity on co-located virtual machines, and they also accomplished keystroke timing attacks [31] to deduce information on the user's input.

b) Model Application: When applying our model to side-channel attacks, almost all entities are involved as shown in Fig. 5. The *provider* configures and chooses the *hardware* and *software* (operating system, hypervisor, etc.) which are supplied by the *manufacturer* and the *developer*, respectively. The input of the *manufacturer* and the *developer* depends on their archetypes. In this scenario it is not reasonable to consider them being *malicious*, but the remaining range from *ostrich* to *defender* may result in input from low quality *hardware* / *software* to specially hardened ones counteracting side-channel attacks. The *provider* also has influence on the feasibility of side-channel attacks, since he configures the system and has to justify his choices of the used *software* and *hardware*.

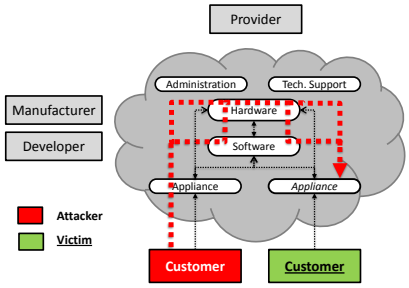


Figure 5. Attacking other customers through side-channels in hardware and/or software.

Similar to the app store scenario, there are two *customers* involved, one is the attacker and one is the victim. As a result from the previous observations, the path of an attacker is to use his appliance to observe characteristics of the *hardware* directly or via the *software* (in this case the underlying physical machine's operating system and especially the hypervisor). Since the attacker needs to create a virtual machine on the same system as the victim, the attacker gains periodic access to the side-channel, i.e., only if his virtual machine is co-located to the victim's machine, he has access. After achieving co-location, the attacker tries to gather information by eavesdropping on the data processed in the attacked appliance of the other involved *customer* (*S1*). Moreover, depending on the information gathered and the infrastructure of the cloud *provider*, the deduced information may allow or ease denial of service attacks (*S10*, *S11*). As already described in the app store scenario, if the attacker is able to steal authentication

information, the confidentiality of data (*S2*) as well as the integrity of the appliance and data is also threatened (*S4*,*S5*), independent of whether it is currently running or stored.

c) Mitigation and Assessment: It is worth mentioning that as a result of these observations the *customer* can do almost nothing to protect himself against side-channel attacks. However, the *customer* can bear additional costs when using physical resources exclusively, which certain *providers* offer. An additional option is using a secured environment like SICE [32] if they are offered by a *provider*. However, if a *provider* is a *defender*, he can monitor *appliance* integrity from the *software* in order to protect his *customers* [33], [34], and even provide recovery options once intrusion has been detected and removed [35].

4) Virtual Machine Escapes:

a) Scenario Description: Ormandy showed that almost all hypervisors contain implementation flaws that could lead to an escape from the virtual machine environment [36]. By escaping the protected environment, the attacker may be able to access the underlying operating system of the physical machine. This way the adversary may be able to attack other virtual machines running on the same physical server with the methods described in the malicious administrator attack scenario. Ormandy especially focused on the most complex parts of the virtual machine hypervisors, which are the instruction subsystem, which handles privileged instructions, and the emulation of I/O devices.

b) Model Application: As shown in Fig. 6 the involved entities are an attacking and a victim *customer* as well as the cloud *provider* and the software *developer*. Similar to the side channel scenario, the cloud *provider* has to configure the system and to choose the used *software* provided by the *developer*. Depending on his skills and motivation, the *developer* of the hypervisor may be in the range from *ostrich* to *stepping stone*, and thus easing or hardening the attacker's task. The attacking *customer* then exploits vulnerabilities in the used hypervisor to break out of his *appliance* and attack another *customer's appliance*. By escaping the *appliance*, the attacker may elevate his access from unprivileged to privileged on the underlying operating system. Depending how extensive his privilege escalation is, the attacked security objectives are analog to those of a malicious administrator, and thus the confidentiality and integrity of the running *appliance* is affected (*S1*, *S4*), as well as of the stored *appliance's* template, because the attacker may gain read or write access on it (*S2*, *S5*) or the network (*S3*, *S6*).

c) Mitigation and Assessment: Although *software* (i.e., hypervisor) is a product of a *developer* and his archetype can determine the safety of a hypervisor, the main responsibility still lies on a *provider*, since he is the one who chooses the *developer* and configures the hypervisor. Thus, a *defender provider* will choose a hardened hypervisor (e.g., Xenon [37]), as well as apply additional security measures like hypervisor integrity check [39]. On the other hand, an *ostrich provider* could also choose secure *software*, but fail to configure it properly or misses to apply security patches when necessary.

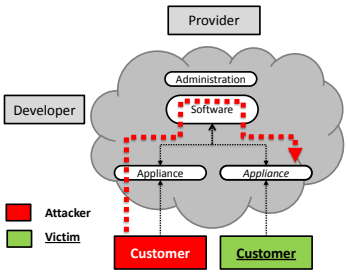


Figure 6. Attacking customer escapes appliance's environment to attack other customers.

B. Constructing What-if Attack Scenarios

Our model is not only useful for describing existing attacks in cloud environments, but also for constructing "what-if" scenarios by combining multiple entities of our model with attacker roles, or by changing an attacker's characteristic. Such what-if attack scenarios derived from our model can lead to possible new attacks which could have been missed in a less-structured assessment of infrastructure cloud security. Cloud customers can use these scenarios to make a security assessment not only based on existing attacks but also potential new attack scenarios. In the following, we demonstrate a subset of what-if attack scenarios based on our model.

1) VM Escape Leading to Large-scale Attacks: In the previous VM escape attacks, a malicious *customer* was attacking other *customers* on the same physical machine. In combination with a *ostrich/charlatan developer* that produces insecure cloud management *software* (e.g., OpenStack[1] misses a large set of security enablements that protect against inside attackers, such as signatures on management commands), the cloud *provider* and *customers* at large can be attacked. For example, by injecting management commands into the insecure management *software*, an attacker can terminate appliances of a large set of *customers*, or consume resources from the *provider* free of charge. Furthermore, if the *manufacturer* of the *hardware* also has the archetype of an *ostrich* or *charlatan*, there may additionally be flaws in the used *hardware*, which could allow an adversary to damage hardware, e.g., by overclocking the central processing unit in an improper way. This would not only lead to additional costs for the *provider*, but probably also to a longer downtime of the concerned physical machines.

2) Insecure Cloud Management Software: To generalize the previous scenario, the security of *cloud management software* has not been studied well enough. For example, vulnerabilities in OpenStack are just beginning to be reported (cf. [40]). Since such *software* will be used by potentially a large set of *providers* vulnerabilities that can be exploited by *cloud customers* will have significant impact, in particular in public cloud offerings.

[1] http://openstack.org

3) Hardware Trojans: Recently Skorobogatov and Woods claim to have discovered a hardware trojan [11]. While this kind of attack has not been seen in a cloud computing scenario, yet, this is a reasonable scenario. In particular, when the *manufacturer* also becomes a *customer* in public clouds that use its *hardware*. By combining the two entities, the malicious *manufacturer* may exploit his one-time access to the hardware later on by using his permanent access to his *appliance*. That way he may be able to steal information from other *customers* or the *provider*. He may also change the way hardware works, threatening the security objectives of availability and integrity not only for other *appliances* but also for the hypervisor and management software.

4) Collusion Attacks in Cloud-of-Clouds: Cloud-of-Clouds systems aggregate multiple clouds in order to tolerate byzantine faults of single clouds. Examples of such systems are presented in [41], [42]. Considering that clouds are operated by different organizations, one may assume that the *administration* and *technical support* of the *providers* do not collude. However, clouds aggregated in a cloud-of-clouds scenario may use the same *software* or *hardware* provided by *malicious/ostrich/charlatan developers* or *manufacturers* respectively, which could form the basis of a colluding attack and diminish the security advantages of cloud-of-clouds systems.

VI. Conclusions and Future Work

We proposed a cloud security threat model that combines a comprehensive system model of infrastructure clouds with a security model focusing on cloud customer security objectives. The threat model differentiates between characteristics and motivations of possible attackers. We applied our model both to the systematic categorization and analysis of existing attacks as well as to the construction of "what-if" attack scenarios based on changing attacker characteristics or combining attackers as they are defined in our model.

By successfully applying the model from a customer's point of view, we showed that it can be used in their security assessment of cloud computing security by providing a better understanding of existing attacks as well as emerging ones. Customers can apply the approach to competing cloud providers, thereby making the services comparable from the perspective of security as a quality attribute. Customers can then choose a service by using approaches such as argumentation logic [48]. This requires that sufficient data about the architecture be available, or that the threat assessment be outsourced to a Trusted Third Party [49].

The model forced us to use a structured approach in describing the attacks, by making us think in terms of entities, components and access rights. The use of the model in a number of scenarios provided us with a number of insights on its usability and generality. Firstly, the model is well-suited for attacks involving technical infrastructure and the behavior of entities, but threats involving governance and compliance, or threats to security monitoring, cannot be easily expressed. These threats depend, respectively, on contractual agreements and the regulatory environment, and the inability of the cloud provider to detect an attack. Neither of these are part of the present version of the model. Secondly, the introduced model proved to be flexible by being able to cover scenarios with multiple instances of the same type. By simply adding another instance of a *provider* it covers the federation of clouds scenario. By considering entities not directly involved in an attack, amplification or reduction of threats by these entities can be investigated.

We consider the following directions as future work for our modeling and analysis efforts. A formalization of our model, such as using process calculi for the system model and utility functions for the attacker goals, may enable an automated and tool-supported security analysis. Furthermore, extending the scope of our model could yield interesting new attack scenarios. For example, we could extend the model to upper abstraction layers in cloud computing, e.g., Platform-as-a-Service, and the consideration of non-technical security threats such as legal or compliance ones (cf. [47]). A systematic categorization and analysis of *protection* mechanisms that counter existing attacks could be beneficial for obtaining a complete picture of attacks and countermeasures in cloud environments, in order to support the cloud customers in their security assessments. In this paper, we only highlighted a subset of possible mitigation strategies.

Acknowledgments

The foundations of this paper were laid in the Dagstuhl seminar on Secure Architectures in the Cloud [50]. This research has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreements number ICT-257243 (TClouds), SEC-261696 (SESAME), and ICT-318003 (TREsPASS), and from the BMBF grant 01lS11008D (SecureClouds) and from the TU Vienna funded HALEY project (Holistic Energy Efficient Management of Hybrid Clouds). This publication reflects only the author's views and the Union is not liable for any use that may be made of the information contained herein.

References

[1] P. Mell and T. Grance, "Effectively and Securely Using the Cloud Computing Paradigm," October 2009.

[2] Cloud Security Alliance, "Top threats to cloud computing v1.0," https://cloudsecurityalliance.org/topthreats/csathreats.v1.0.pdf, 2010.

[3] ENISA, "Cloud Computing Risk Assessment," ENISA, Tech. Rep., 2009.

[4] B. Hay, K. Nance, and M. Bishop, "Storm Clouds Rising: Security Challenges for IaaS Cloud Computing," in *Proceedings of the 2011 44th Hawaii International Conference on System Sciences*, ser. HICSS '11. Washington, DC, USA: IEEE Computer Society, 2011, pp. 1–7.

[5] W. Pieters, "Security and privacy in the clouds: a bird's eye view," in *Computers, Privacy and Data Protection: an Element of Choice*, S. Gutwirth, Y. Poullet, P. De Hert, and R. Leenes, Eds. Dordrecht: Springer, 2011, pp. 445–457.

[6] F. Rocha and M. Correia, "Lucy in the sky without diamonds: Stealing confidential data in the cloud," in *Proceedings of the 1st International Workshop on Dependability of Clouds, Data Centers and Virtual Computing Environments (DCDV, with DSN'11)*, June 2011.

[7] T. Ristenpart, E. Tromer, H. Shacham, and S. Savage, "Hey, You, Get Off of My Cloud: Exploring Information Leakage in Third-Party Compute Clouds," in *CCS '09: Proceedings of the 16th ACM conference on Computer and Communications Security*. New York, NY, USA: ACM, 2009, pp. 199–212.

[8] Amazon Web Services, "Summary of the Amazon EC2 and Amazon RDS Service Disruption in the US East Region," http://aws.amazon.com/message/65648/, April 2011.

[9] S. Bugiel, S. Nürnberger, T. Pöppelmann, A.-R. Sadeghi, and T. Schneider, "Amazonia: when elasticity snaps back," in *Proceedings of the 18th ACM conference on Computer and communications security*, ser. CCS '11. New York, NY, USA: ACM, 2011, pp. 389–400.

[10] P. Mell and T. Grance, "The NIST definition of cloud computing," Special Pulication 800-145, September 2011.

[11] S. Skorobogatov and C. Woods, "Breakthrough silicon scanning discovers backdoor in military chip," in *CHES*, 2012, pp. 23–40.

[12] R. Winther, O.-A. Johnsen, and B. Gran, "Security assessments of safety critical systems using HAZOPs," in *Computer Safety, Reliability and Security*, ser. Lecture Notes in Computer Science, U. Voges, Ed. Springer Berlin / Heidelberg, 2001, vol. 2187, pp. 14–24.

[13] M. Bishop and C. Gates, "Defining the insider threat," in *Proceedings of the 4th annual workshop on Cyber security and information intelligence research*, ser. CSIIRW '08. New York, NY, USA: ACM, 2008, pp. 15:1–15:3.

[14] E. LeMay, M. Ford, K. Keefe, W. Sanders, and C. Muehrcke, "Model-based security metrics using adversary view security evaluation (ADVISE)," in *Quantitative Evaluation of Systems (QEST), 2011 Eighth International Conference on*, sept. 2011, pp. 191–200.

[15] S. E. Schechter, "Toward econometric models of the security risk from remote attack," *IEEE Security and Privacy*, vol. 3, pp. 40–44, 2005.

[16] W. Pieters, S. H. G. Van der Ven, and C. W. Probst, "A move in the security measurement stalemate: Elo-style ratings to quantify vulnerability," in *NSPW '12: Proceedings of the 2012 New security paradigms workshop*. ACM, 18-21 Sep 2012, forthcoming.

[17] R. Brunette, G. Mogull, "Security guidance for critical areas of focus in cloud computing," *Cloud Security Alliance*, 2010.

[18] D. Catteddu and G. Hogben, "Cloud computing risk assessment," *European Network and Information Security Agency (ENISA)*, 2009.

[19] Deloitte, "Cloud security risk map," http://tinyurl.com/935ktap, 2012.

[20] J. Oberheide, E. Cooke, and F. Jahanian, "Exploiting Live Virtual Machine Migration," in *BlackHat DC Briefings*, Washington DC, February 2008.

[21] T. Garfinkel, B. Pfaff, J. Chow, M. Rosenblum, and D. Boneh, "Terra: A Virtual Machine-Based Platform for Trusted Computing," *SIGOPS Oper. Syst. Rev.*, vol. 37, no. 5, pp. 193–206, 2003.

[22] F. Zhang, J. Chen, H. Chen, and B. Zang, "Cloudvisor: retrofitting protection of virtual machines in multi-tenant cloud with nested virtualization," in *23rd ACM Symposium on Operating Systems Principles (SOSP'11)*. ACM, 2011.

[23] S. Bleikertz, A. Kurmus, Z. A. Nagy, and M. Schunter, "Secure cloud maintenance - protecting workloads against insider attacks," in *7th ACM Symposium on Information, Computer and Communications Security (ASIACCS'12)*. ACM, May 2012.

[24] C. Gentry, "Fully homomorphic encryption using ideal lattices," in *41st annual ACM symposium on Theory of Computing*. ACM, 2009.

[25] M. Van Dijk and A. Juels, "On the impossibility of cryptography alone for privacy-preserving cloud computing," in *5th USENIX conference on Hot topics in security (HotSec'10)*. USENIX, 2010.

[26] S. Potter, S. M. Bellovin, and J. Nieh, "Two-Person Control Administration: Preventing Administration Faults Through Duplication," in *Proceedings of the 23rd conference on Large installation system administration*, ser. LISA'09. Berkeley, CA, USA: USENIX Association, 2009, pp. 15–27.

[27] J. Wei, X. Zhang, G. Ammons, V. Bala, and P. Ning, "Managing security of virtual machine images in a cloud environment," in *Proceedings of the 2009 ACM workshop on Cloud computing security*, ser. CCSW '09. New York, NY, USA: ACM, 2009, pp. 91–96.

[28] Amazon Web Services, "Amazon elastic compute cloud user guide," http://awsdocs.s3.amazonaws.com/EC2/latest/ec2-ug.pdf, June 2012.

[29] M. Balduzzi, J. Zaddach, D. Balzarotti, E. Kirda, and S. Loureiro, "A Security Analysis of Amazon's Elastic Compute Cloud Service," in *Proceedings of the 27th Annual ACM Symposium on Applied Computing*, ser. SAC '12. New York, NY, USA: ACM, 2012, pp. 1427–1434.

[30] W. Zhou, P. Ning, X. Zhang, G. Ammons, R. Wang, and V. Bala, "Always up-to-date: scalable offline patching of vm images in a compute cloud," in *Proceedings of the 26th Annual Computer Security Applications Conference*, ser. ACSAC '10. New York, NY, USA: ACM, 2010, pp. 377–386.

[31] D. X. Song, D. Wagner, and X. Tian, "Timing analysis of keystrokes and timing attacks on ssh," in *USENIX Security Symposium*, D. S. Wallach, Ed. USENIX, 2001.

[32] A. M. Azab, P. Ning, and X. Zhang, "SICE: a hardware-level strongly isolated computing environment for x86 multi-core platforms," in *Proceedings of the 18th ACM conference on Computer and communications security*, ser. CCS '11. New York, NY, USA: ACM, 2011, pp. 375–388.

[33] A. M. Azab, P. Ning, E. C. Sezer, and X. Zhang, "HIMA: A Hypervisor-Based Integrity Measurement Agent," in *Proceedings of the 2009 Annual Computer Security Applications Conference*, ser. ACSAC '09. Washington, DC, USA: IEEE Computer Society, 2009, pp. 461–470.

[34] T. Garfinkel and M. Rosenblum, "A virtual machine introspection based architecture for intrusion detection," in *In Proc. Network and Distributed Systems Security Symposium*, 2003, pp. 191–206.

[35] M. Kirkpatrick, G. Ghinita, and E. Bertino, "Resilient authenticated execution of critical applications in untrusted environments," *Dependable and Secure Computing, IEEE Transactions on*, vol. 9, no. 4, pp. 597 –609, july-aug. 2012.

[36] T. Ormandy, "An Empirical Study into the Security Exposure of Hosts of Hostile Virtualized Environments," Google, Inc., Tech. Rep., Feb 2007, http://taviso.decsystem.org/virtsec.pdf.

[37] J. McDermott and L. Freitas, "A formal security policy for xenon," in *Proceedings of the 6th ACM workshop on Formal methods in security engineering*, ser. FMSE '08. New York, NY, USA: ACM, 2008, pp. 43–52.

[38] R. Sailer, T. Jaeger, E. Valdez, R. Caceres, R. Perez, S. Berger, J. L. Griffin, and L. v. Doorn, "Building a MAC-Based Security Architecture for the Xen Open-Source Hypervisor," in *Proceedings of the 21st Annual Computer Security Applications Conference*, ser. ACSAC '05. Washington, DC, USA: IEEE Computer Society, 2005, pp. 276–285.

[39] A. M. Azab, P. Ning, Z. Wang, X. Jiang, X. Zhang, and N. C. Skalsky, "Hypersentry: enabling stealthy in-context measurement of hypervisor integrity," in *Proceedings of the 17th ACM conference on Computer and communications security*, ser. CCS '10. New York, NY, USA: ACM, 2010, pp. 38–49.

[40] NIST, "National vulnerability database, vulnerability summary for CVE-2012-2654," http://web.nvd.nist.gov/view/vuln/detail?vulnId=CVE-2012-2654, June 2012.

[41] C. Basescu, C. Cachin, I. Eyal, R. Haas, A. Sorniotti, M. Vukolic, and I. Zachevsky, "Robust data sharing with key-value stores," in *42nd Annual IEEE/IFIP International Conference on Dependable Systems and Networks*, Boston, MA, USA, June 2012.

[42] A. Bessani, M. Correia, B. Quaresma, F. André, and P. Sousa, "DepSky: Dependable and Secure Storage in a Cloud-of-Clouds," in *Proceedings of the sixth conference on Computer systems*, ser. EuroSys '11. New York, NY, USA: ACM, 2011, pp. 31–46.

[43] I. M. Abbadi, C. Namiluko, and A. Martin, "Insiders analysis in cloud computing focusing on home healthcare system," in *The 6th International Conference for Internet Technology and Secured Transactions (ICITST-2011)*. IEEE, Dec. 2011, pp. 350–357.

[44] B. Grobauer, T. Walloschek, and E. Stocker, "Understanding cloud computing vulnerabilities," *Security Privacy, IEEE*, vol. 9, no. 2, pp. 50 –57, march-april 2011.

[45] T. Garfinkel and M. Rosenblum, "When Virtual is Harder than Real: Security Challenges in Virtual Machine Based Computing Environments," in *HOTOS'05: Proceedings of the 10th conference on Hot Topics in Operating Systems*. Berkeley, CA, USA: USENIX Association, 2005, pp. 20–20.

[46] A. Behl, "Emerging security challenges in cloud computing: An insight to cloud security challenges and their mitigation," in *Information and Communication Technologies (WICT), 2011 World Congress on*, dec. 2011, pp. 217 –222.

[47] D. Molnar and S. Schechter, "Self hosting vs. cloud hosting: Accounting for the security impact of hosting in the cloud," in *Proceedings of the Ninth Workshop on the Economics of Information Security (WEIS)*, Jun. 2010.

[48] J. Rowe, K. Levitt, S. Parsons, E. Sklar, A. Applebaum, and S. Jalal, "Argumentation logic to assist in security administration," in *NSPW '12: Proceedings of the 2012 New security paradigms workshop*. ACM, 18-21 Sep 2012, forthcoming.

[49] C. W. Probst, M. A. Sasse, W. Pieters, T. Dimkov, E. Luysterborg, and M. Arnaud, "Privacy penetration testing: How to establish trust in your cloud provider," in *European Data Protection: In Good Health?*, S. Gutwirth, R. Leenes, P. De Hert, and Y. Poullet, Eds. Springer Netherlands, 2012, pp. 251–265.

[50] S. De Capitani di Vimercati, W. Pieters, and C. W. Probst, "Secure architectures in the cloud," Dagstuhl, Germany, Technical Report 11492, December 2011.

B.2 Elicitation of Requirements for an inter-organizational Platform to Support Security Management Decisions

Julian Dax, Benedikt Ley, Sebastian Pape, Christopher Schmitz, Volkmar Pipek, and Kai Rannenberg. Elicitation of requirements for an inter-organizational platform to support security management decisions. In *10th International Symposium on Human Aspects of Information Security & Assurance, HAISA 2016 ,Frankfurt, Germany, July 19-21, 2016, Proceedings.*, 2016. URL `https://www.cscan.org/openaccess/?paperid=295`

Proceedings of the Tenth International Symposium on
Human Aspects of Information Security & Assurance (HAISA 2016)

Elicitation of Requirements for an inter-organizational Platform to Support Security Management Decisions

J. Dax[2], B. Ley[2], S. Pape[1],C. Schmitz[1], V. Pipek[2] and K. Rannenberg[1]
[1] Goethe University Frankfurt, Chair of Mobile Business & Multilateral Security, Germany
[2] University of Siegen, Institute of Information Systems, Germany
e-mail: {julian.dax; benedikt.ley; volkmar.pipek}@uni-siegen.de
{sebastian.pape; christopher.schmitz; kai.rannenberg}@m-chair.de

Abstract

Due to new regulations in Germany energy providers are required to obtain IT security certificates. Especially small and medium-sized energy providers struggle to fulfill these new requirements. Since most of them are in the same situation, we are dealing with the question on how to support their collaboration using a web-based platform. We elicited criteria from energy providers on how such a platform should be designed to support them. The main contribution is a set of requirements for the collaboration platform along with the implications for its implementation. The focus of this work is not on technical innovation but on how existing technologies and best practices can be adopted for the needs of small and medium-sized energy providers.

Keywords

Usable Security, Security Management, Security Assessment, Security Perception

1. Introduction

The European Program for Critical Infrastructure Protection (EPCIP) was recently implemented in national laws in Germany. The IT security law requires providers of critical infrastructures to get certifications for their security. This especially concerns energy providers as they also have to comply with industry-sector-specific regulations laid out in the Energy Industry Act (EnWG). There is no de minimis rule if the definition for critical infrastructure is fulfilled. As a consequence, in particular small and medium-sized energy providers struggle to fulfill the requirements. Compared to larger providers, they have the handicap that there is a low budget for IT security and that no experts for IT security are employed there. One of their first challenges in order to meet the criteria is to introduce an information security management system (ISMS). Most of the providers mainly do this to comply with the new regulation. When the ISMS is put to work, the energy providers should make use of it to monitor and improve the IT security of their systems.

Most of the energy providers are uncertain how to start and may need to hire external consultants to support them. The aim of the project SIDATE is to support them to continuously improve their security. Since many of the small and medium-sized energy providers face very similar challenges, a natural solution to support them is to stimulate inter-organizational collaboration. This should be done by building an

inter-organizational collaboration platform for energy providers. The platform should enable the energy providers to share their knowledge about IT security in a structured way.

In this paper, we describe the requirements elicitation process with the energy providers. We aimed to engage them very early in the design process. It showed that many of the criteria are not domain-specific for energy providers. Therefore, we believe that other domains can profit from those criteria as well. Our contribution is a set of requirements for the collaboration platform along with the implications for its construction.

The remainder of this paper is organized as follows: Section 2 discusses related work. Section 3 describes the used methodology. Section 4 sketches the results of the first workshop with the energy providers. The planned modules for our collaboration platform are shown in Sect. 5. In Sect. 6, we describe the design criteria for the collaboration platform collected from energy providers.

2. Related Work

2.1. Collaboration platforms and expertise sharing

The “endeavor to understand the nature and characteristics of cooperative work with the objective of designing adequate computer-based technologies.“ (Bannon & Schmidt 1989) has always been the aim of Computer Supported Cooperative Work (CSCW). Therefore, collaboration platforms have been a major field of research in CSCW. Inside this field, the aspect of inter-organizational needs for such platforms can be studied. While ‘inter-organizational information systems’ (IOIS) are automated information systems shared by two or more organizations (Cash & Konsynski 1985), CSCW applications provide “capabilities beyond simple information access to facilitate communication and collaboration among partners” (Drury & Scholtz 2005). The term ‘knowledge sharing’ is used for artifact-centered studies, while the communication-centered ‘expertise sharing’ focuses on the actor (Ackerman et al. 2013). Further, expertise sharing focuses on the “self-organized activities of the organization’s members and emphasizes the human aspects” (Ackerman et al. 2013). There have been a number of studies of expertise sharing in CSCW in different fields of application: For example, Doherty et al. (Doherty et al. 2012) studied inter-organizational coordination mechanisms in software development and Hobson et al. (Hobson et al. 2011) studied the information sharing needs and practices in municipal governments. Bharosa et al., (Bharosa et al. 2010) conducted a study on multi-agency disaster response and identified the problem that “actual level of information sharing across different organizations is often limited, although it is being promoted”. For energy providers the German association of municipal corporations "Verband kommunaler Unternehmen" (VKU) offers an efficiency comparison/benchmark, but unfortunately no online platform is offered.

Proceedings of the Tenth International Symposium on
Human Aspects of Information Security & Assurance (HAISA 2016)

2.2. Shared Risk Analysis, ISMS and Stakeholders' Engagement

Karlsson et al. (Karlsson et al. 2015) regard ISMS to manage information systems in inter-organizational collaborations. The difference to our use-case is, that the energy providers do not collaborate in the sense of sharing business processes. The reason for them to use our collaboration platform would be that they face the same challenges and are able to exchange experiences. Faily (Faily 2014) reports on engaging stakeholders in the design of a secure system. Our platform also aims to engage the stakeholders; not on the system itself but rather on sharing experience and expertise on how to design secure systems.

When it comes to implementing information security policies in organizations, Arif (Arif 2011) studied five factors which determine the willingness to comply with these policies: culture, awareness, training, risk perception and re-enforcement. In his study, the cultural factor was the most impactful. Reichard et al. (Reichard et al. 2011) studied barriers to the successful implementation of such policies and how to overcome them. Like Arif, they stress the importance of a "security culture" in the organization. Moreover, they stress the need for collaborative implementation of such policies. Another related factor in the successful introduction of IT-security policies identified by Reichard et al. is that the principles and benefits of IT-security have to be communicated and "sold" to the organization.

Apart from that, in the US the concept of Information Sharing Analysis Centers (ISACs) can be found. Those non-profit organizations gather and analyze IT security-related information within critical infrastructure sectors (e.g. electricity) and provide analysis results, security strategies and general information to their members. In contrast to that, our approach focuses more on the individual assessing and benchmarking of the energy provider's security level (ISAC Council 2004).

3. Methodology

In order to elicit the target group-specific requirements, three two-hour workshops with different stakeholder groups were conducted. In total, eleven experts from eight energy providers attended the workshops. Most participants were IT security officers or IT managers from energy providers, but also representatives from national interest groups were present.

Seven experts from six different energy providers attended the first workshop. After an introductory talk by the organizer, each of the attendees introduced themselves based on a short questionnaire which addressed, for instance, general characteristics of their company and their experience in IT security. Afterwards, the experts were invited to discuss the platform's requirements and their expectations in a moderated discussion.

The workshop's results were subsequently discussed in an additionally internal design workshop, where eight members from the project partners were involved. As a result, several mockups visualizing the platform's functionalities were sketched.

In another workshop, five experts from six energy providers attended as well as three employees from two interest groups. After the mockups had been presented, the discussion which was moderated by using the card-technique, was opened. The participants were asked to formulate the platform's must-have and nice-to-have requirements on different colored cards. After 10 minutes, the cards were collected and sorted in content-related clusters on a pin board. Then, all cards were discussed in an open discussion.

4. Energy Providers' Needs

Before we started to design our platform, we collected the energy providers' requirements for a collaboration platform. Our assumption was that for the communication between the energy providers, a web-based solution which allows asynchronous communication is most helpful. Mainly, because there is no need to install additional software which lowers the threshold to participate. This was confirmed by the energy providers during the workshop. The following modules were considered helpful by the energy providers: a wiki, a forum, a questions and answers module, a glossary, training modules for further education for security officers and other employees, checklists, a place to exchange documents, benchmarks, security assessment modules and a general module to support the launch of an ISMS.

5. A Platform Supporting Security Management

From the results of the first workshop with the energy providers, we inferred that the most relevant modules for the energy providers which should be implemented in the 1st iteration are:

- A security assessment module, which allows the energy providers to get feedback about their security level.
- A security measures module, which provides information and recommendation to energy providers about measures which they can implement in order to strengthen their IT-security.
- A question and answer module.

All modules should allow the energy providers to give feedback and exchange their experiences. We describe them below:

5.1. Security Assessment Module

The security assessment module follows a questionnaire-based quantitative methodology (Frangopoulos et al. 2014). The module allows energy providers to perform a self-assessment in order to assess and to improve their current IT security level. This is done by answering an online questionnaire which is provided on the proposed platform (see figure 1). The answers of other energy providers to these questions are also shown in aggregated form in order to allow the user to compare

his/her organization to others. Additionally, the best rated questions asked by other community members related to the current topic are also shown.

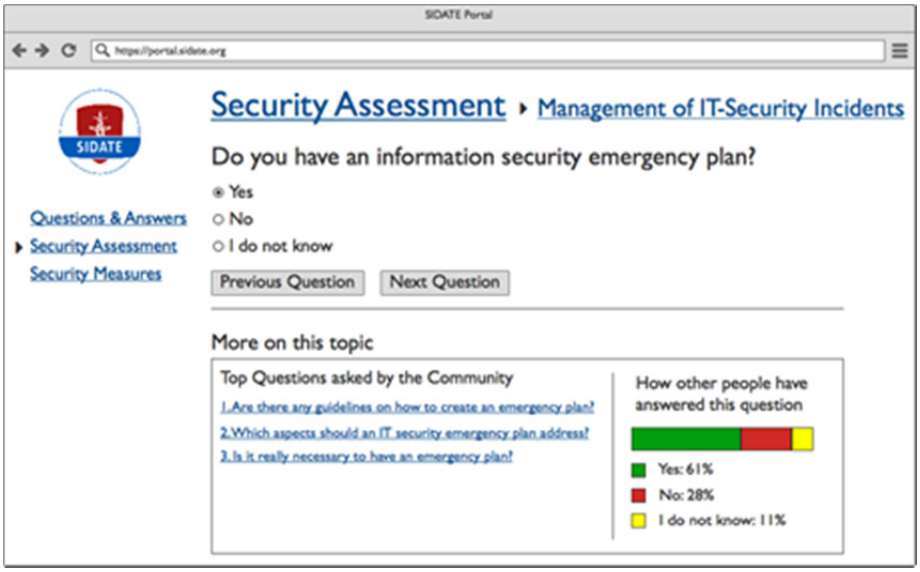


Figure 1: Mockup of the Security Assessment Module

5.2. Question and Answer Module

In the questions and answers module registered users can ask questions related to IT-security. These questions can be categorized by tags and be assigned to ISO/IEC 27002 controls.

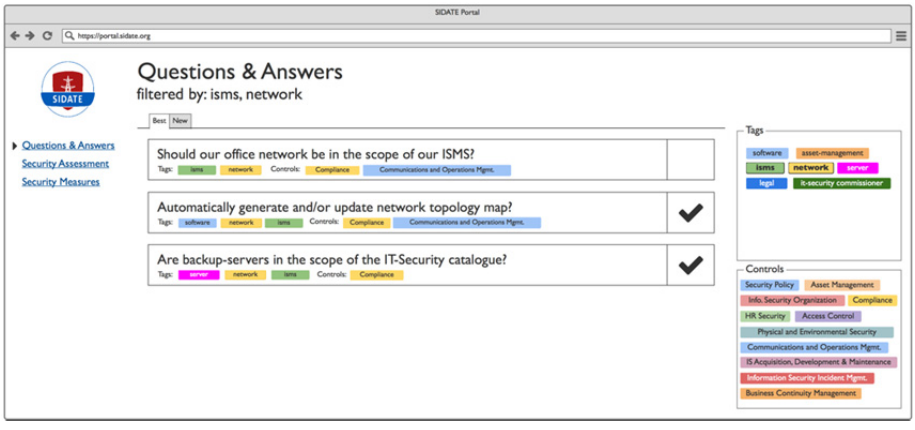


Figure 2: Mockup of the Questions and Answers Module

A side bar on the right (see Figure 2) allows users to select these tags and controls to filter the questions. Questions can be answered by other users, and answers can be marked as correct by the user who posted the question. Additionally, questions and answers can be rated and either sorted by rating or creation date.

5.3. Security Measures Module

The security measures module is a catalogue of security measures, which is maintained by security experts. Each security measure is categorized by one or more tags and assigned to one or more specific ISO/IEC 27002 controls. Users can comment on the measures and rate them according to their costs, efficacy and usability.

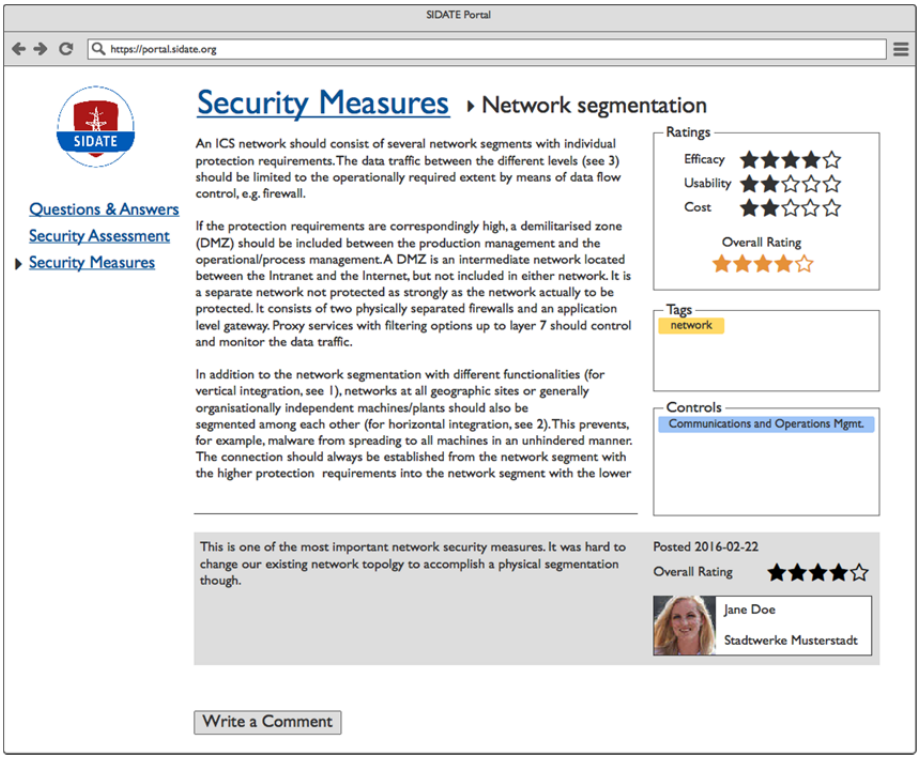


Figure 3: Mockup of the Security Measures Module

6. Elicitation of Criteria for the Fundamental Platform Design

In the second workshop with the energy providers, we presented the created mockups to the participants to show the possible functionality of the proposed platform. Then we asked them to write down mandatory and nice-to-have requirements the platform has to fulfil to be usable for them. We got 28 individual answers that we could cluster into four major categories: *(1) platform members, (2)*

Proceedings of the Tenth International Symposium on
Human Aspects of Information Security & Assurance (HAISA 2016)

confidentially/data privacy, (3) integration into exiting workflows, (4) general usability of the platform. After we had clustered the participants' answers, we discussed each category to expose the motivations behind the requirements and initial approaches to solution.

6.1. Platform Participants and Data Privacy

The categories *platform members* and *confidentially/data privacy* were discussed together because of several overlaps between both categories. As expected, we could determine that participants had essential concerns about the privacy in respect to sensitive IT-security related data they would share across the platform. However, these concerns basically did not refer to the platform itself or its operator but to other platform members.

While it seems to be acceptable to share information with other energy providers, respectively their employees, participants were worried about the participation of external experts like information security consultants or lawyers. Even if they see an advantage in the qualified and skilled feedback from such persons, we discovered two significant concerns we have to deal with. (1) External experts could misuse the platform for advertising purposes and could flood energy providers with personalized offers based on the platform content. (2) Non-reliable platform members could use the visible content and questions by individual energy providers to identify and make use of possible security flaws.

Based on these initial insights, we developed and discussed several approaches with the workshop participants in order to find possible solutions that protect the energy providers' data and identity on the one hand and make use of the expertise from third parties on the other hand. While some of the approaches that are listed below are mutually exclusive, others complement each other.

- It is necessary that the platform supports **restricted and moderated access** for new members. Individuals or organizations that intend to participate to the platform need to be validated by the platform operator and have to agree to suitable terms of use in order to get access.
- Different UI views based on the user's organization and role could be used to **anonymize individuals and organizations** to external experts. While energy providers are able to see each other's questions, answers and other activities, other participants can only see the content but not the corresponding author. Energy providers should be able to rate the experts' contributions in order to improve their reputation. Instead of getting unwanted advertising, the energy providers can now proactively inquire consultancy service based on the experts' reputation.
- Instead of giving experts access to the platform, energy provides should be able to mark their contribution as *expert approved.* This means that the contribution rests on the result from consultancy service or legal advice the respectively user made use of before. This approach completely excludes

third parties from the platform and only allows the **indirect passing of expert's assessments and opinions** via the energy providers.
- As reliable organizations, the **interest groups for energy providers could undertake the role of experts** on the platform and contribute to energy providers' questions. However, the participating representatives of the interest groups in the workshop made clear that they do not have profound expertise to give sufficient answers to all questions. The only practicable approach is that they inform about legal changes and regulations on information security for energy provider.

6.2. Integration into Existing Workflows

The aim of the platform is to support participating energy providers to improve their information security and fulfill legal regulations. Thus, another important topic we have discussed with the workshop participants was that the effort they have to put into using the platform must not exceed the potential benefit. Several requirements given by the participants dealt with the question on how can the platform and its functionality be integrated into users' existing workflows.

- As a result from the self-assessment module the platform should provide individual **checklists and tools** that help the users' implementing required information security measures. In a first step this should predominantly aim at the fulfillment of statutory provisions (in case of energy providers in Germany the implementation of an ISMS according to ISO/IEC 27001).
- The self-assessment should also contribute to **internal information security audits**, e.g. the regular validation of measures and processes.
- It should be possible to **export results** from self-assessment to reuse them for internal reports (e.g. to be presented to the management) or other processes and workflows like the information security related controlling.

6.3. General Usability of the Platform

The remaining requirements that came up during the workshop focused on the general usability and will only be described briefly here because of their generality. Essentially the participants expect that the content on the platform is well-structured and maintained. There should be a moderator who leads discussions to an outcome, ensures that new topics/questions are created in the right section and prevents duplicates. Also the platform has to be up to date and deprecated content needs to be marked as such.

7. Conclusion and Future Work

Due to new regulatory requirements for critical infrastructures, especially small and medium-sized energy providers struggle to get their IT security certified. Because they face very similar challenges, we proposed a new concept for a collaboration platform in order support them to collaboratively improve their IT security.

To elicit the specific requirements of how such a platform should be designed, we conducted workshops with different stakeholder groups. As a result, we identified a set of functions and requirements which the platform has to fulfill.

There are three elementary modules. A central role plays the security assessment module for assessing and benchmarking the energy provider's security level. The second module is the security measures module which describes the most relevant IT security measures including the practical experiences by other energy providers. Finally, there is the questions and answers module which allows them to share their experiences with both other energy providers as well as with external experts.

Because the platform processes highly sensitive data, aspects in regard to data privacy have a very high priority for the stakeholders. This includes, for instance, having different UI views to anonymize individuals and organizations to external experts, and having a restricted and moderated access for new members. Also the integration into existing workflows plays a central role. For example the self-assessment should provide individual checklists and tools according to the ISO/IEC 27001 and should contribute to the internal information security audit. Besides that, the general usability of the platform was mentioned as essential requirement.

The next step is to implement the proposed concept and to iteratively refine the platform's functions based on user feedback. As future work, it would be interesting to analyse to what extent the platform can be transferred to other domains.

8. Acknowledgement

This research was developed in the context of the project SIDATE which is funded by the German Federal Ministry of Education and Research (BMBF) within its funding priority ”IT Security for Critical Infrastructures". Grant number: 16KIS0239K, 16KIS0240.

9. References

Ackerman, M.S. et al., 2013. Sharing Knowledge and Expertise: The CSCW View of Knowledge Management. *Computer Supported Cooperative Work (CSCW)*, 22(4-6), pp.531–573.

Arif, M., 2011. What Matters Most Among Human Factors to Comply With Organisation's Information Security Policy? In 5th International Symposium on Human Aspects of Information Security and Assurance, HAISA 2011, London, UK, July 7-8, 2011. Proceedings. pp. 35–46.

Bannon, L.J. & Schmidt, K., 1989. CSCW - Four Characters in Search of a Context. *DAIMI Report Series*, 18(289).

Bharosa, N., Lee, J. & Janssen, M., 2010. Challenges and obstacles in sharing and coordinating information during multi-agency disaster response: Propositions from field exercises. *Information Systems Frontiers*, 12(1), pp.49–65.

Cash, J.I. & Konsynski, B.R., 1985. IS Redraws Competitive Boundaries. *Harvard Business Review*, 63, pp.134–142.

Doherty, G., Karamanis, N. & Luz, S., 2012. Collaboration in Translation: The Impact of Increased Reach on Cross-organisational Work. *Computer Supported Cooperative Work (CSCW)*, 21(6), pp.525–554.

Drury, J. & Scholtz, J., Evaluating Inter-Organizational Information Systems. In *Inter-Organizational Information Systems in the Internet Age*. Inter-Organizational Information Systems in the Internet Age, pp. 266–296.

Faily, S., 2014. Engaging Stakeholders in Security Design: An Assumption-Driven Approach. In Eighth International Symposium on Human Aspects of Information Security & Assurance, HAISA 2014 ,Plymouth, UK, July 8-9, 2014. Proceedings. pp. 21–29.

Frangopoulos, E.D., Eloff, M.M. & Venter, L.M., 2014. Human Aspects of Information Assurance: A Questionnaire-based Quantitative Approach to Assessment. In Eighth International Symposium on Human Aspects of Information Security & Assurance, HAISA 2014 ,Plymouth, UK, July 8-9, 2014. Proceedings. pp. 217–229.

Hobson, S.F. et al., 2011. Towards Interoperability in Municipal Government: A Study of Information Sharing Practices. In *Human-Computer Interaction – INTERACT 2011*. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 233–247.

ISAC Council, 2004. *A Functional Model for Critical Infrastructure Information Sharing and Analysis*, White Paper (31 January).

Karlsson, F. et al., 2015. Inter-Organisational Information Sharing - Between a Rock and a Hard Place. *HAISA*, pp.71–81.

Reichard, A., Quirchmayr, G. & Wills, C.C., 2011. Challenges in Implementing Information Security Policies. In 5th International Symposium on Human Aspects of Information Security and Assurance, HAISA 2011, London, UK, July 7-8, 2011. Proceedings. pp. 22–34.

B.3 Easing the Burden of Security Self-Assessments

Christopher Schmitz, Andre Sekula, Sebastian Pape, Volkmar Pipek, and Kai Rannenberg. Easing the burden of security self-assessments. In *12th International Symposium on Human Aspects of Information Security & Assurance, HAISA 2018, Dundee, Scotland, August 29-31, 2018, Proceedings.*, 2018

Proceedings of the Twelfth International Symposium on
Human Aspects of Information Security & Assurance (HAISA 2018)

Easing the Burden of Security Self-Assessments

C. Schmitz[1], A. Sekulla[2], S. Pape[1], V. Pipek[2] and K. Rannenberg[1]

[1] Goethe University Frankfurt, Chair of Mobile Business & Multilateral Security, Germany
[2] University of Siegen, Institute of Information Systems, Germany
e-mail: {christopher.schmitz; sebastian.pape; kai.rannenberg}@m-chair.de;
{andre.sekulla; volkmar.pipek}@uni-siegen.de

Abstract

A web-based platform was developed to support the inter-organisational collaboration between small and medium-sized energy providers. Since critical infrastructures are subject to new security regulations in Germany, the platform particularly serves for the exchange of experience and for mutual support in information security. The focus of this work is the security self-assessment component. In order to ease the burden of going through a long questionnaire we have implemented small, motivating modules that are spread across the platform. The data entered is used for an individual risk assessment but also for a fine granular inter-organisational security benchmarking which builds a common added value for the entire community on the platform and strengthens the community building process. We implemented a prototype of the platform and evaluated the it in a focus group.

Keywords

Security Management, Security Self-Assessment, Collaborative Knowledge Management

1 Introduction

Gathering information for risk and security self-assessments can be a cumbersome task. In general, the security managers need to answer an often long collection of questions built on established standards (e.g. Swanson 2001, ISO/IEC 27019, IEC 62443). For instance, the NIST security self-assessment contains more than 200 questions (Swanson 2001). Self-assessments offer advantages over external security audits: they are less expensive, they can be implemented in local organisational routines, and they allow more control on critical information about an organisations' IT infrastructure. But they are also challenging: the actors' bias towards the inner-organisational discourses may leave blind spots. Furthermore, analyses, as well as decisions for counter-measures, require a continuous improvement of competencies with regard to existing as well as future IT infrastructures and the related threats. These challenges are particularly relevant for small and medium-sized enterprises (SMEs) that provide infrastructural services, and which often do not have the capacities to run a full-fledged information security department and rely on external expertise (Dax et al. 2017).

In many areas, individuals and organisations with a local lack of expertise turn to support communities on the internet. These communities are not only valuable in offering their members concrete support to solve a specific problem, they also offer an interaction space to collaboratively consolidate and improve the general knowledge on the issues at stake, and offer additional problem solving strategies (e.g. by means of recommender systems, cf. Ackerman et al. 2013). This approach cannot immediately be transferred to areas with specific vulnerabilities, e.g. information security in power grid infrastructures. Framing conditions like the high sensitivity of the infrastructure-related information, legal or regulatory requirements, and the complexity of dependencies between grid technologies, IT systems supporting their management, and possible threats require a more cautious approach to unlock the helpful dynamics of community processes.

We have developed a platform for security managers supporting small and medium-sized energy providers. The central tool of this platform is a self-assessment component to support security managers to manage the recent legal requirements to monitor and improve the information security of their infrastructures. In our approach, users can model the existing information security measures of their infrastructure (in terms of security controls following ISO/IEC 27001) using security maturity levels, which can then be compared and published in an anonymised way to the results from other participating organisations. The platform then provides information (in a Q&A section) on improving with regard to specific controls, as well as a controlled community section in which strategies of improvement can be discussed with other information security managers. We built small modules which are shown in other parts of the platform. Those modules allow the users to answer the questions or update the maturity levels along the way when interacting with other parts of the platform. By making use of motivational elements and showing questions one by one in other parts of the platform, we aim to ease the burden of going through a lengthy list of questions. This is especially the case when respondents update the answers entered and need to decide if the current answer is still valid. Lessons from other platforms showed, that structured processes of information consolidation and improvement through users help the perceived value of the information provided dramatically.

The remainder of this paper is organised as follows: Section 2 discusses related work, Section 3 gives an overview of our platform, and Section 4 discusses the connection of self-assessment with user motivation and community building. Section 5 reports about a brief evaluation. Section 6 concludes and outlines future research.

2 Related Work

With the World Wide Web as a breakthrough technology, building knowledge communities became an actual practice in professional contexts (e.g. Lesser et al. 2000). Although these community platforms intended an open, flexible support for problem-solving processes, the delicacy of the social and business-related processes behind the “innocent” knowledge exchange very soon became apparent: Articulating a problem was often considered as uncovering a personal or organisational deficit, solutions that were offered came with unclear quality assurances, and the work of

narrowing down a problem as well as developing a solution that would fit all local needs went far beyond simple "Q&A" patterns (Pipek and Won, 2003).

For platforms hosting knowledge communities, several strategies were developed to ease these problems. The idea of "FAQ" (Frequently asked questions) developed to relieve experts from answering the same basic questions over and over. It was combined with processes to keep them up to date (e.g. the "Answer Garden" system, Ackerman and McDonald 1996). Pipek and Won (2003) suggested to focus more on connecting users looking for a problem solution with experts who could help them, less on making knowledge explicit and store it online. For particularly sensitive issues, the anonymity of the person asking for help as well as of persons answering is guaranteed (e.g. patientslikeme.com).

Self-assessment as another technique to counter negative effects of "deficit disclosure", and even allows a continuous monitoring and improvement, has become a heavily discussed approach in learning communities (e.g. Castle and McGuire, 2010). To some extent, self-assessment approaches also help in organisational learning (e.g. in the general improvement of IT infrastructures, e.g. Curley 2004, in approaches of quality management, e.g. Saunders and Mann 2005, and – with regard to information security – e.g. Swanson 2001). But this was never done in combination with online support for knowledge communities. There exist the so-called "Information Sharing Analysis Centres" (ISACs). ISACs are organisations that gather and analyse security-related information from their members and provide them with analysis results and reports. In contrast to them our approach addresses the individual organisation and provides them with individual risk analysis and benchmarking scores. Furthermore, our platform enables a direct knowledge sharing.

3 The SIDATE Platform

Especially SMEs often struggle to achieve an adequate security level, although some of them are obliged to get certified against the ISO/IEC 27001. This holds for instance for energy providers and other critical infrastructures in Germany. A natural solution to support them is to stimulate collaboration. For this, we have built an inter-organisational collaboration platform for energy providers. It enables energy providers to assess their security level and to improve their security also by inter-organisational discussions. We systematically elicited the requirements in several workshops (Dax et al. 2016). The platform consists of four main components aiming to support knowledge sharing between the organisations:

- **Security measures catalogue:** The security measures component is a catalogue of security measures which is maintained by security experts. Users can comment on the measures, suggest new measures and rate them according to their costs, efficacy and usability.
- **Questions and answers:** The Q&A component should support and structure inter-organisational discussions. Registered users can ask security-related questions and can finally mark answers as correct. All users can rate questions and answers and can either sort them by rating or creation date. In order to have a

more structured inter-organisational communication threads can be filtered according to tags or security controls.

- **Document sharing:** In the document sharing component the participating organisations can share relevant documents in a structured way, e.g. best practices or official documents specifying the binding legal requirements.
- **Security self-assessment:** The security self-assessment component constitutes the core component of the platform. Using this component, organisations can assess their security risk level in order to better understand their exposure to relevant security risks. Moreover, they can compare their security status (on different abstraction level) with that of similar organisations.

In the following, we focus on the self-assessment component which constitutes the central element of the platform. It consists of the three sections data input, benchmarking and risk assessment that are complemented by three superordinate modules being spread across the platform. We describe them below:

3.1 Data Input Section

The first step of the risk assessment process is to gather the required information. The necessary user data is entered in the data input section (see Fig. 1). The organisations model the security measures of their infrastructure by assessing the maturity levels of the implemented security controls (in terms of controls following the ISO/IEC 27001). Here, the widely known ISO/IEC 27019 security controls (which are more specific security controls for the energy utility industry) are used as questionnaire items. Since they equally address technical and organisational aspects of information security they represent a wide range of security measures that can be implemented in an organisation. The items are structured in the same categories and sub-categories the security managers already know from the original standard. The users are furthermore supported by the feature to show either all controls, only those controls that are not assessed yet or only those controls that have already been assessed which makes sense in order to check in a user-friendly way whether all controls are still up to date.

3.2 Benchmarking Section

The benchmarking section (see Fig. 2) enables organisations to compare their security status with similar organisations. Their maturity levels are juxtaposed (in an anonymised way) with that of other organisations.

Proceedings of the Twelfth International Symposium on
Human Aspects of Information Security & Assurance (HAISA 2018)

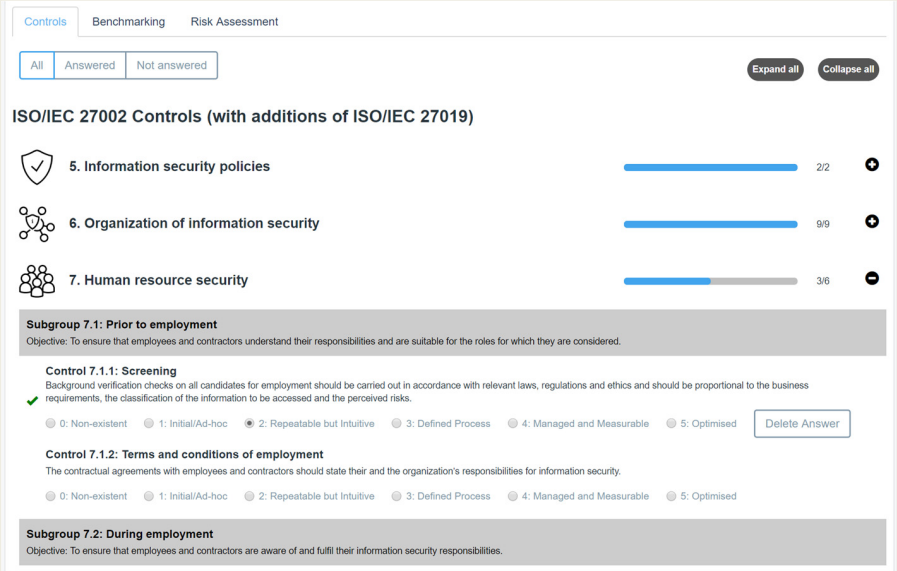


Figure 1: Data Input Section

For each control, the organisation's maturity level is shown along with the average maturity level by the other organisations. For a more in-depth analysis, the distribution of maturity levels per control is also presented as well as a relative benchmarking score which indicates how well the organisation performs compared to the others. In this section, one can also re-assess the maturity levels. The benchmark is shown on different abstraction levels: on a control level and on the aggregated levels of the control groups and sub-groups of the ISO/IEC 27019. The groups and sub-groups are presented in the same structure as in the original standards, like in the data input section.

3.3 Risk Assessment Section

In the risk assessment section a scenario-based risk analysis is conducted to calculate the organisation's security risk score as well as the risk for a collection of relevant attack scenarios. This supports the security managers in identifying the most critical risks they are exposed to. Describing the risk assessment framework and the other data sources would go beyond the scope of this work.

3.4 Superordinate Modules

Additionally, we have implemented three superordinate modules directly supporting the self-assessment component. The modules are displayed in other components of the platform aiming to connect the different parts of the platform in order to stimulate the users to frequently assess respectively to re-assess security controls.

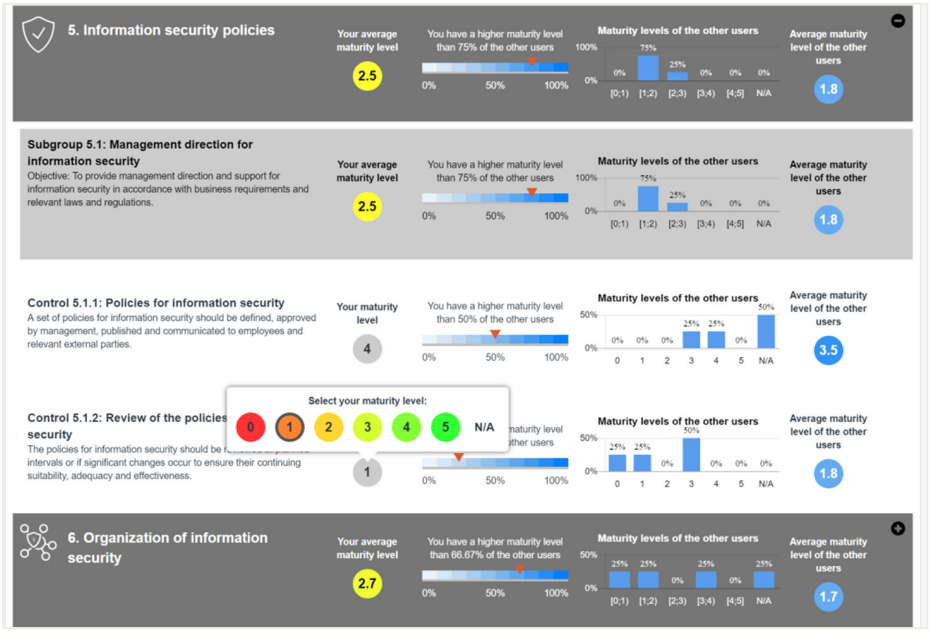


Figure 2: Benchmarking Section

Figures 3 and 4 show their graphical use interfaces. By requesting to keep the data complete and up-to-date we try to keep the entire data on a representative level.

1.) A control that has already been evaluated may have an obsolete maturity level and should be updated to obtain a more representative status. Therefore, the first module (see Fig. 3) requests the user to update resp. to re-assess a security control at regular intervals. This is also important from the perspective of information security management systems, since they require constant and iterative handling of information security measures.

2.) In case of missing maturity levels the second module requests the user to evaluate the security controls that have not been evaluated yet. In particular, the aim is to ensure that the data is complete. The more controls have been evaluated, the better the outcomes of the risk assessment and the better they can be compared with other results. The presented controls are further prioritized with regard to their information value for the risk assessment, e.g. to enable a new attack scenario in the risk assessment. The module also indicates such information.

3.) The third module, shown in Figure 4, is positioned in the security measures catalogue. While a user is viewing such a measure in detail, he or she gets asked to evaluate the respective security control for the self-assessment component. Again, this should improve the data completeness and up-to-dateness.

Proceedings of the Twelfth International Symposium on
Human Aspects of Information Security & Assurance (HAISA 2018)

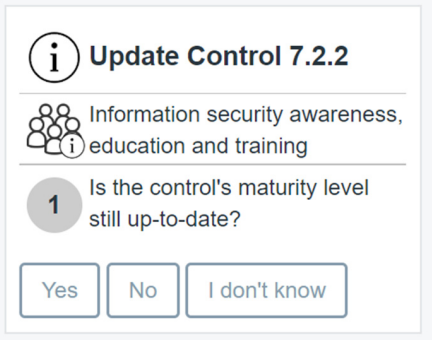


Figure 3: Update Control Module

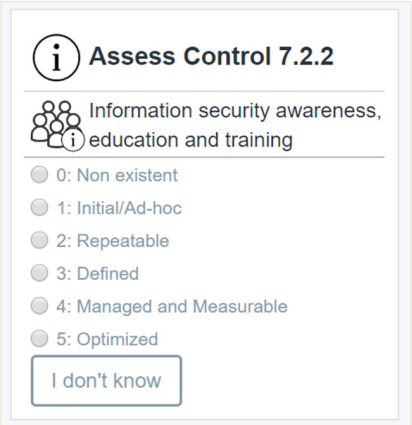


Figure 4: Assess Control Module

4 Usability Aspects

To improve interaction and activity in the SIDATE platform, the interaction between users need to be carefully planned. Looking at the individual user, a good usability and interesting collaboration anchors need to be provided. But it is also important to have the further development of the associated community in mind.

4.1 Motivating Updates and Additional Input

One way to increase activity on the online platform is to keep the entry barriers as low as possible (Girgensohn and Lee, 2002). The self-assessment tool serves as a guided entry to model the maturity level of an organisation's IT security. Later changes can be easily made as soon as a user is logged in on the platform without the questionnaire. He can easily add further data and information to his information security status without having to navigate directly into the associated self-assessment module in order to additionally reach the subordinate category in such a way that he can evaluate the corresponding control.

Section 3 described the self-assessment component in more detail. The component does not include any community functions itself. Since this component may contain sensitive data, functions for exchange and interaction between users could be counterproductive. They could lead to falsified data or no input of the requested data being carried out. The modules presented below are primarily intended to ensure that the dataset entered is complete and up-to-date. This enables the self-assessment and the benchmarking to work properly on these data and make meaningful comparisons. Only after the own data has been entered, the other users' ratings become visible as a direct comparison. This should again increase the motivation to enter complete data.

The asses control module (Fig. 4) indicates that when the corresponding control is evaluated, a new attack scenario is activated within the risk analysis of the self-assessment module. This should increase the motivation to enter complete ratings and unlock a kind of success because "individuals are more likely to gain self-based

achievement rather than enjoyment in the process of sharing knowledge" (Yang and Lai, 2010). Hence, while users are viewing such a measure in detail they get asked to evaluate the respective security control for the self-assessment module and the benchmarking process. Again, this should improve the up-to-dateness and data completeness and is implemented through the related control module.

4.2 Supporting the Community Building Process

The activity of the users of a platform is an important aspect of the community building process. Beside the user activity, another goal of an online platform for cooperation is the creation of added value for all parties involved. Girgensohn and Lee (2002) describe the so-called socio-technical-capital as "a resource produced as a side effect of technology-mediated social interaction". Resnick (2001) notes that it can be accumulated and made available to create value for people. It should influence the users among themselves in such a way that they interact more with each other. To encourage users to participate further, it is recommended to "repeat social interaction" (Kollock, 1996) which is implemented in particular with the help of additional modules directly related to the presented self-assessment module. It is intended to encourage users to constantly interact with the platform. The self-assessment itself has no functions for direct interaction between the users but the small modules have indirect effects on further interactions on the entire platform, as they allow for an anonymous comparison with the results others have provided.

If there is a need for an improvement in their own information security landscape, users can start to enter and participate in online discussions that are specific to the controls where deficits may be rooted in. It is not necessary to disclose that there are deficits in a user's own organisation but the discussions can aim for a general optimization with regard to that control. It remains (formally) open whether a participant is looking for or providing expertise – this positioning is left to the discourse itself. The aim is to awake the interest to exchange ideas with other users of the platform in order to learn from their experiences and to profit from the resulting social-technical-capital. Thus, with the help of the self-assessment module and the associated superordinate modules, a community building process is initiated that increases the activity of all interaction methods integrated on the platform.

5 Evaluating the Platform in a Focus Group

To evaluate the platform, we have conducted a workshop with ten experts from eight small or medium-sized energy providers. Due to the legal requirements, the majority of the organisations were certified against ISO/IEC 27001 so they successfully went through all the necessary processes. Therefore, most of the participants had good security know-how. One of them was a trainer for ISO/IEC 27001 security auditors.

We have presented the most relevant platform features in a live demo. The attendees could always interrupt the presentation and ask questions to make sure they understood everything. Afterwards, the experts were invited to discuss the platform in a moderated discussion. We asked them for general feedback and for suggestions for improvement

Proceedings of the Twelfth International Symposium on
Human Aspects of Information Security & Assurance (HAISA 2018)

based on their own experiences. We also stimulated discussions among the experts and moderated it in the way to work out the most relevant aspects.

The participants emphasised the simple structure and the user-friendly design of the platform. Their comments and the way they discussed the platform and its functions also clearly demonstrated they understood the purpose of the different functions and how to use them. Apart from those usability aspects, many of the comments were addressing the ISO/IEC 27001 certification. There was consensus among the experts that the platform was helpful for an internal pre-audit before the official ISO/IEC 27001 audit starts. They argued for instance that the organisations have to conduct a risk analysis prior to the official audit anyway, and such a self-assessment would be very helpful for SMEs who often struggle to identify and assess the risks they are exposed to. The experts also agreed that the approach to go through the ISO/IEC 27019 controls makes a lot of sense because this is what the auditor finally checks.

The users' positive evaluations on both the platform's usability and its general ideas have a positive effect on the users' activity and it strengthens the community building process which helps the entire community. To further improve the platform the experts suggested integrating a recommender feature that derives optimal security measures and recommends a list of actions to the security team. According to the benchmarking component, it would be useful to have a benchmarking with companies already certified against ISO/IEC 27001.

6 Conclusion and Future Work

Security self-assessment frameworks support security managers to assess their organisation's security level. Applying those frameworks can be a cumbersome task since many of them are based on long questionnaires. Apart from that, additional information and inter-organisational discussions, e.g. with regard to the selection of security measures, can often be helpful especially for SMEs who often do not have the capacities to run a full-fledged security department. In order to address these issues, a web-based collaboration platform for security management was developed, supporting energy providers. The security self-assessment component constitutes the central feature of the platform. It helps security managers to identify relevant attack scenarios and allows them to benchmark their security status with that of similar organisations. Complementarily, small modules were implemented that are spread across the platform. They allow the users to complete or update the data needed for the self-assessment along the way when interacting with other parts of the platform. By making use of motivational elements and showing questions one by one in other parts of the platform, we aim to ease the burden of security self-assessments (e.g. going through a long questionnaire).

Furthermore, we have implemented a prototype of the platform and have evaluated it in a focus group, concentrating on usability aspects but also on the conceptual ideas of the platform. The next steps are to address the experts' feedback and to work on a recommender function for security measures based on the results of the security risk analysis. Another open task is to analyse how to design the inter-organisational sharing of recommended measures in a privacy preserving way.

7 Acknowledgement

This research was supported by the German Federal Ministry of Education and Research (grant numbers: 16KIS0239K, 16KIS0240). We thank Leon Alexander Herrmann and David Bug for their contribution to the prototype implementation.

8 References

Ackermann, M. S., McDonald, D. W. (1996), „Answer Garden 2: Merging Organizational Memory with Collaborative Help", *CSCW'96*, ACM Press, pp97-105.

Castle, S. R. and McGuire, C. (2010), "An analysis of student self-assessment of online, blended, and face-to-face learning environments: Implications for sustainable education delivery", *International Education Studies*, Vol. 3, No. 3, p36.

Curley, M. G. (2004), Managing information technology for business value: practical strategies for IT and business managers (IT best practices series), Intel press.

Dax, J., Ivan, A., Ley, B., Pape, S., Pipek, V., Rannenberg, K., Schmitz, C. and Sekulla, A. (2017): "IT Security Status of German Energy Providers", Technical Report, Cornell University, arXiv.

Dax, J., Ley, B., Pape, S., Schmitz, C., Pipek, V. and Rannenberg, K. (2016): Elicitation of Requirements for an inter-organizational Platform to Support Security Management Decisions, *10th Int. Symposium on Human Aspects of Information Security & Assurance*, HAISA 2016, Frankfurt, Germany, Proceedings.

Girgensohn, A., Lee, A. (2002), "Making Web Sites Be Places for Social Interaction", *CSCW'02*, New Orleans, Louisiana, USA.

Kollock, P. (1996), "Design Principles for Online Communities", *Harvard Conference on the Internet and Society,* Cambridge, MA.

Lesser, E. L., Fontaine, M. A. and Slusher, J. A. (eds.) (2000), *Knowledge and Communities.* Butterworth-Heinemann, Oxford, UK.

Pipek, V. and Won, M. (2002), "Communication-oriented Computesr Support for Knowledge Management", *Informatik/Informatique - Magazine of the Swiss Informatics Societies*, Vol. 1, pp39–43.

Resnick, P. (2001), "Beyond Bowling Together: SocioTechnical Capital", *J.M. Carrol (ed.)*, *Human-Computer Interaction in the New Millennium, Addison-Wesley*, pp647-672.

Saunders, M. and Mann, R. (2005), "Self-assessment in a multi-organisational network", *IJQRM*, Vol. 22, Issue 6, pp554-571.

Swanson, M. (2001), "Security Self-Assessment Guide for Information Technology Systems", *NIST Special Publication 800-26.*

Yang, H.-L. and Lai, C.-Y. (2010), "Motivations of Wikipedia content contributors", *Computers in Human Behavior*, Vol. 26, Issue 6, pp1377-1383.

B.4 A structured comparison of the corporate information security

Michael Schmid and Sebastian Pape. A structured comparison of the corporate information security. In *ICT Systems Security and Privacy Protection - 34th IFIP TC 11 International Conference, SEC 2019, Lisbon, Portugal, June 25-27, 2019, Proceedings*, pages 223–237, 06 2019. doi: 10.1007/978-3-030-22312-0_16. URL https://doi.org/10.1007/978-3-030-22312-0_16

A Structured Comparison of the Corporate Information Security Maturity Level

Michael Schmid[1,2(✉)] and Sebastian Pape[1,3(✉)]

[1] Chair of Mobile Business & Multilateral Security, Goethe University Frankfurt, Frankfurt, Germany
{michael.schmid,sebastian.pape}@m-chair.de
[2] Hubert Burda Media Holding KG, Munich, Germany
[3] Chair of Information Systems, University of Regensburg, Regensburg, Germany

Abstract. Generally, measuring the information security maturity is the first step to build a knowledge information security management system in an organization. Unfortunately, it is not possible to measure information security directly. Thus, in order to get an estimate, one has to find reliable measurements. One way to assess information security is by applying a maturity model and assess the level of controls. This does not need to be equivalent to the level of security. Nevertheless, evaluating the level of information security maturity in companies has been a major challenge for years. Although many studies have been conducted to address these challenges, there is still a lack of research to properly analyze these assessments. The primary objective of this study is to show how to use the analytic hierarchy process (AHP) to compare the information security controls' level of maturity within an industry in order to rank different companies. To validate the approach of this study, we used real information security data from a large international media and technology company.

Keywords: Information security · Information security management · ISO 27001 · Analytic hierarchy process · Information security controls · Capability maturity model · Security maturity model · Security metrics framework

1 Introduction

Information security can only be measured indirectly [6]; unfortunately there is still no gold standard. One way to indirectly measure it is to use metrics and KPIs [1] which aim to approximate the real status of information security. This approach is not always reliable [22]. Some information to build those metrics are obtained from technical systems (e.g. firewalls, intrusion detection/prevention systems, security appliances). However, most of these metrics and KPIs have to be quantified by humans and are therefore prone to errors.

Published by Springer Nature Switzerland AG 2019
G. Dhillon et al. (Eds.): SEC 2019, IFIP AICT 562, pp. 223–237, 2019.
https://doi.org/10.1007/978-3-030-22312-0_16

This can lead to possible inaccuracies, measurement errors, misinterpretations, etc. [4]. If these metrics are then compared across the board, the information security managers face a major challenge. As a consequence, this could lead to bad decisions based on wrong conclusions. Moreover, by just comparing the metrics, without any weighting the specifics of the respective industry are not considered. Thus, a prioritisation within the comparison is not possible [3]. This problem is reinforced when the comparison of information security metrics between different companies or departments would take place [9], which is exactly on of the current challenges enterprises face today: How to compare their (sub-)companies of a specific industry (e.g. eCommerce) in terms of information security.

The main goal of this paper is to compare the effect of multiple factors in the information security assessment process. Aiming at achieving this goal, the analytic hierarchy process (AHP) is applied. The Analytical Hierarchy Process (AHP) is one of the most commonly used Multiple Criteria Decision Methods (MCDM), combining subjective and personal preferences in the information security assessment process [20]. It allows a structured comparison of the information security maturity level of companies with respect to an industry [25] and to obtain a ranking [13]. This allows us to define a separate weighting of information security metrics for each industry with to respect their specifics while using a standardized approach based on the maturity levels of the ISO 27001:2013 controls [12]. ISO 27001 was in particular selected, because this standard is shown to be mature, widespread and globally recognized. This minimizes the additional effort for collecting the required metrics. In this study, the maturity level is based on a hierarchical, multi-level model to analyze the information security gap for the ISO 27001:2013 security standard [20]. As a prerequisite for the comparison, we assume companies have implemented an information security management system (ISMS) in accordance with ISO 27001 [26].

To validate the approach of this study, we used real information security data (i.e. security controls' maturity level) from Hubert Burda Media (HBM) a large international media and technology company consisting of over 200 individual companies. This provides sufficient data with a high degree of detail in the area of information security. The result from our AHP-based approach is then compared with the perceived status of information security by experts.

The remainder of this work is structured as follows: In Sect. 2 we give a brief overview of related work. Section 3 describes our methodology when we developed our approach shown in Sect. 4. Our results are shown in Sect. 5 followed by a discussion and our conclusion in Sect. 6, respectively Sect. 7.

2 Background and Related Work

In addition to the differences in the assessment of information security, all assessment procedures have in common that the ratings of the maturity level and the weighting of weights remain separate judgements and are not allocated to a common overall value in the sense of an 'information security score'. It is therefore

up to the evaluator to carry out the respective evaluation, as he or she is forced to choose between these two quantitative aspects of the evaluation, i.e. the ratings on the one hand and the weighting on the other [15]. In contrast to this, the works of Boehme [6] and Anderson [3] deal more with the economic impact of investments in information security. The focus of this work is to compare the degree of maturity within an industry. This could later lead to a monetary assessment of information security or maturity.

A solution which involves to merging ratings and weights and thus integrates different assessment measures at the same time offers multi-attribute decision-making procedures [8]. These are methods that offer support in complex decision-making situations, i.e. when a decision has to be made in favour of one of several options against the background of several decision criteria (so-called attributes).

The prerequisite for using the multi attribute decision procedure is, as described above, the determination of weights. A popular method of doing this is the Analytic Hierarchy Process (AHP) method developed by Saaty [23]. Nasser [2] describes how to measure the degree of maturity using AHP. In contrast to our paper which deals with the comparison of the maturity level within an industry, Nasser [20] focuses on the determination of inaccurate expert comparison judgement in the application of AHP.

Some recent works deals with this problem setting using the AHP but there exist further restrictions. Watkins [27] uses for his approach not the control maturity level and is only valid in the cyber security environment. Bodins' [5] approach is based on the comparison of the CIA-Triangle and not on ISO 27001-controls. Peters [21] has already shown the application of AHP in the domain of project management but did not use real data to validate the approach.

2.1 Multiple Criteria Decision Methods

Multi criteria decision problems which could be solved with a multiple-criteria decision analysis method (MCDM) are a class of procedures for the analysis of decision or action possibilities characterized by the fact that they do not use a single superordinate criterion, but a multitude of different criteria. Problems in evaluating multiple criteria consist of a limited number of alternatives that are explicitly known at the beginning of the solution process. For multiple criteria, design problems (multiple objective mathematical programming problems), the alternatives are not explicitly known. An alternative (solution) can be found by solving a mathematical model. However, both types of problems are considered as a kind of subclass of multi-criteria decision problems [17]. MCDM helps to determine the best solution from multiple alternatives, which may be in conflict with each other. There are several methodologies for MCDM such as: Analytical hierarchical process (AHP), Grey relational analysis (GRA), Technique for order preference by similarity to ideal solution (TOPSIS), Superiority and inferiority ranking (SIR), Simple additive weighting (SAW), and Operational competitiveness rating (OCRA) [7].

2.2 The Analytical Hierarchy Process

The AHP, is a method developed by the mathematician Saaty [24] to support decision-making processes. Because of its ability to comprehensively analyse a problem constellation in all its dependencies, the AHP is called 'analytical'. It is called a 'process' because it specifies how decisions are structured and analysed. In principle, this procedure is always the same, which makes the AHP an easy-to-use decision tool that can be used more than once and is similar to a routine treatment [16]. The goal of the Analytic Hierarchy Process method is to structure and simplify complex decision problems by means of a hierarchical analysis process in order to make a rational decision. The AHP breaks down a complex evaluation problem into manageable sub-problems.

3 Research Methodology

Many companies use the maturity level measurement of the controls from ISO standard 27001 to obtain a valid and reliable metric. The ISO standard is well established and the maturity assessment of the standard's controls is an adequate possibility to create a picture of the information security processes of a company. While this might be sufficient for a continuous improvement within the same company, a problem arises if one wants to compare the information security processes of different companies or departments. Depending on the field of industry, some of the processes might be more important than others.

The general aim of this approach is to determine which company within an industry is better or worse in a (sub)area of information security, in order to create transparency among the companies within an industry concerning information security. Positive effects of this approach would be the improvement or deterioration of the information security in a sector within an industry recognizable up to the question where the management should invest money economically for information security in order to improve a sector.

We define the requirements in the next subsection, then determine the proper algorithm and finally describe the data collection for our approach.

3.1 Requirements

The most important requirement is that the metrics we rely on should be easy to gather. Assuming that the investigated company is running an information security management system (ISMS), a natural approach is to rely on the controls of the ISO/IEC 27001 standard and their maturity level. Existing data (e.g. information security maturity level) should be used wherever possible. Furthermore, the approach should consider the environment of the industry in which the company is located. Additionally, the information gathering should be repeatable and stable. Comparing and evaluating over a long period should be possible as well as an overall as an comparison of security levels of business units or companies in a similar area. Finally, the approach should allow it to visualize and explain the results of the comparison and allow to derive the areas where companies could improve.

3.2 Algorithm Selection

Taking all requirements into account, our problem is a multi-dimensional decision problem, and thus can be addressed by a multiple-criteria decision analysis method (MCDM). Our comparison criteria (dimensions) are the ISO/IEC 27001 controls and we compare the different companies based on their corresponding maturity levels for each control. Thus, the MCDM needs discrete, quantitative input and a criteria weighting method. Since the underlying controls are hierarchically and therefore very structured, the chosen method/model should reflect that also.

This leads us to the analytical hierarchy process (AHP) as a best fit method in the above described context. The AHP is a mature structured technique for organizing and analyzing complex decisions, combining subjective and personal preferences. The AHP has been the most widely used technique of multi-criteria decision making during the last twenty five years [19]. The advantage of this method over the utility value analysis, for example, is that it goes beyond the evaluation of ideas and generates a clear selection recommendation. Its hierarchical structuring of decision making fits well to the ISO/IEC 27001 controls' hierarchy and the qualitative evaluation part of the AHP is very much in line with the maturity level for information security. Since the AHP compares the maturity level for each control company-wise, it naturally allows to understand where each company's security level is ranking related to each control. Additionally, the weight of each criteria (control) can be easily derived. In the concrete application case it is possible to compare the importance of individual controls of ISO 27001 very granularly with each other (pairwise). This is in particular necessary in order to be able to establish an industry reference. Furthermore, the AHP enables precise calculations of weights, in this case the information security maturity ratings of companies in a specific sector.

Thus, we used a paired comparison questionnaire based on the AHP to compare controls and their maturity level for an industry.

3.3 Data Collection

To test the above approach it is necessary to set up the model and verify it with real data. We need a maturity assessment of the ISO/IEC controls and to weight them according to the considered industry. We focused on the eCommerce industry for the following reasons:

- Available data from a large range of companies
- Excellent data quality and validity
- High actuality of the existing data
- Very good know-how available in the expert assessment of the industry.

Maturity Assessment of ISO/IEC 27001 Controls. We collected data from Hubert Burda Media (HBM), an international media and technology company (over 10,000 employees, more than 2 billion annual sales, represented in

over 20 countries). This group is divided into several business units that serve various business areas (including print magazines, online portals, e-commerce, etc.). The business units consists of over 200 individual companies with about 30 of them being in the eCommerce industry. Each subsidiary operates independently of the parent corporation. There is a profit center structure, so the group acts as a company for entrepreneurs and the managing directors have the freedom to invest money into information security or choose the appropriated level of security.

We will briefly describe how this data is collected before going into more detail on the data used for the comparison. Each individual company in the group operates its own Information Security Management System (ISMS) in accordance with ISO/IEC 27001:2013, which is managed by an Information Security Officer (ISO) on site and managed by a central unit in the holding company. As part of the evaluation of the ISMS, the maturity level for the respective ISO 27001 controls is ascertained - very granularly at the asset level. The maturity level is collected/updated regular once a year as part of a follow-up.

First, the information values of the respective company (e. g. source code, customer data, payment data, etc.) are determined according to the protection goals of confidentiality, integrity and availability and assigned to a technical system (e. g. application, client, server, etc.).

Second, these technical systems undergo a threat analysis[1] of the assets in relation to the respective asset type as part of information security risk management. The threat analysis is classically evaluated with regard impact[2] and the probability of occurrence. This results in an aggregated risk value (1–5) for each asset after a pre-defined settlement. This risk value is later transferred to the control valuation as the *target maturity level*. In this way, a comparison is made between the protection requirements of the information values and the protection level of the respective (IT) system.

Third, the control evaluation is then carried out using the Cobit maturity level. The controls are dynamically selected[3] according to the previously evaluated threats. The Cobit maturity level is a 6-step evaluation scale (0–5) with which a continuous improvement can be measured and a potential improvement can be identified. This allows it to evaluate the actual maturity level per control and asset. The assessment of the current status of the controls is carried out by the information security officer of the respective company. The collected data is therefore not technical data but subjectively quantified data with a possible bias. Although, the evaluated data is reviewed by further experts, a complete review cannot be carried out due to resource limits. The target maturity level is already determined by the risk value/protection level of the system. This provides a clear picture of the ISMS status at a very granular asset level.

[1] Threat catalogue according to ISO/IEC 27005:2011.
[2] Referring to the protection goals of confidentiality, integrity and availability.
[3] By a predefined threat/control matrix.

Fourth, the picture is completed by the Cobit maturity analysis of the IT-/ISM processes[4]. For each of these processes, the controls (e. g. A.16 for incident management) are evaluated with an actual maturity level [10]. In the later evaluation (typically by means of a spider graphic) the complete ISO 27001 standard is evaluated with the aid of the Cobit degree of maturity [14].

The available data is very granular on asset level (application, client, server, etc.). However, although the companies are from the same industry, they do not necessarily have the same kind of assets. Thus, we decided to abstract from the assets and to aggregate the data at company level. To do this automatically, we used the mean value of all evaluated assets per control. For the following proof of concept, we only show data from 5 companies.

4 The Approach - the AHP-Implementation

In this section, we discuss how the AHP is applied to our comparison. The first step of the AHP, to model the problem as a decision hierarchy, we have already done by deciding that our decision-criteria will be the ISO/IEC 27001 controls. The goal is clearly defined: to find the subsidiary within the company with the best information security/level of maturity within an industry. Appendix A of ISO 27001 helps us to select criteria and sub criteria, which is divided into 14 Control Categories, 35 Control Objectives and 114 Controls (see Fig. 1).

The next step is the prioritization of all criteria and sub criteria (Sect. 4.1). This represents the domain specific part of the AHP calculations and it only needs to be done once per domain. It is followed by the evaluation of the alternatives (Sect. 4.2). The alternatives represent the agile part of the calculation. We describe in the corresponding section, how the evaluation can be directly derived from the maturity level of a company's control. Based on the individual evaluations and prioritizations of controls, the AHP uses a mathematical model to determine a precise weighting of all alternatives in relation to the respective criteria and assembles them in a percentage order (Sect. 4.3).

In the next subsections we describe in detail how the AHP was used and show how the applied AHP model was implemented in a statistical software (in this case in R).

4.1 Pairwise Comparison of the Control Categories and Controls

The characteristics of an industry have a significant influence on the pairwise comparison when comparing the individual controls. If the information security of companies is to be compared with each other, e.g. in the e-commerce sector, it will differ significantly from that of companies in other sectors, e.g. publishing or the manufacturing industry. On the one hand this is due to the different business models within the industries, because the IT strategy and the information

[4] Business Continuity Management, Compliance, Incident Management, Information Security Management, Organizational Information Security, Protection Requirement Assessment.

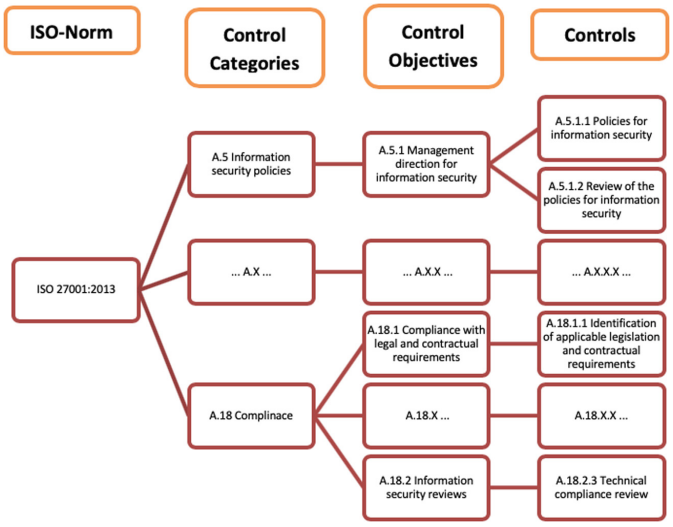


Fig. 1. Exemplary ISO 27001 Appendix A structure

security strategy are derived from the business strategy. On the other hand this is due to the different focus in information security. For example, the eCommerce industry is very focused on application development and (confidentiality) protection of customer data, whereas the highest commodity to be protected in the manufacturing industry is the availability of systems.

The decision-maker must compare each criterion with its pair and denotes which of the two criteria appears more important to him/her. This method of pairwise comparisons allows the decision-maker to elicit a very precise evaluation from the multitude of competing criteria. The comparisons must be carried out specifically for one industry (e. g. eCommerce). In the case of our hierarchy based on the ISO/IEC 27001 controls, 91 pairwise comparisons have to be made for the control categories and 208 for the controls, respectively. This leads to a ranking order in which the criteria are ranked according to their importance.

The comparison is done as follows: Each result of a pairwise comparison of two criteria entered in the evaluation matrix shows how much more significant a criteria is in relation to the criteria of the level above. To do this, refer to the scale in Table 1a. In order to make a comparison for one criteria, i.e. the control categories, we compare the individual control categories with each other. The authors made this comparison in a straight forward Excel spreadsheet. The assessment of the relative importance of the criteria at the criterion level can be found in Table 1b. These pairwise comparisons are always carried out by an expert with the background knowledge and with reference to the industry (here eCommerce). The comparison for the sub criteria, the controls, follows the same guidelines.

Table 1. AHP scores and their application

AHP Score	Verbal description
9 8	Extreme preference
7 6	Very strong preference
5 4	Strong preference
3 2	Moderate preference
1	Equal preference

(a) *Fundamental AHP Score*

Sub criteria A	Sub criteria B	A/B	Score
Control A.12.1.1[1]	Control A.12.1.2	B	$\frac{1}{7}$
Control A.12.1.1	Control A.12.1.3	B	$\frac{1}{7}$
Control A.12.1.1	Control A.12.1.4[4]	B	$\frac{1}{7}$
Control A.12.1.2[2]	Control A.12.1.3	B	$\frac{1}{3}$
Control A.12.1.2	Control A.12.1.4	A	3
Control A.12.1.3[3]	Control A.12.1.4	A	3

[1]Documented operating procedures [2]Change management
[3]Capacity management [4]Separation of development

(b) *AHP Comparison with sub criteria (Controls) from control group A.12.1*

4.2 Pairwise Evaluation of the Controls' Maturity Levels

The alternatives in our example are the information security maturity of 5 eCommerce companies of HBM. For each control and each company there is a corresponding maturity level based on the Cobit Maturity Model. 0 represents the worst and 5 the best result, always in relation to the evaluation of a control. As already discussed in Sect. 3.3, the maturity levels for each company were based on assets and we aggregated the maturity levels by calculating the average maturity level for each control over all evaluated assets of the respective company.

For the pairwise comparison, the gap between the comparative maturity levels of two companies' controls is considered to decide which company is doing better at a specific control. For that purpose, we need to map the 6-stage scale of the Cobit maturity grade gaps (see Table 2) to the 9-stage AHP score. The result is a table where each GAP Cobit interval represents an AHP score, which is verbally described. An exemplary calculation can be found in Table 2c). Alternative A (Company 1) is compared with the alternatives B (Company 2 to 5). A Cobit GAP -2 (i.e. 1–3) means hat Company 2 is 2 control maturity better than Company 1, the AHP score is, corresponding to the Cobit GAP interval, 4, respectively 1/4. This can be used to calculate which of the 5 companies performs best in Control A.5.1.1.

The step of comparing the companies' maturity levels for each control represents the business unit specific part of the analysis. Note that, due to our mapping of the GAP Cobit interval and the AHP score, this can be done fully automatic if the corresponding maturity levels are provided. The pairwise comparison, the calculation of the difference and the 'translation' to the GAP intervals is done in the statistics software R.

Table 2. Combined GAP of Cobit Maturity Model and AHP score

Cobit Maturity Model	Cobit level
Optimized	5
Managed and Measurable	4
Defined Process	3
Repeatable but Intuitive	2
Initial/Ad Hoc	1
Non-existent	0

(a) *Maturity Model vs. level*

AHP Score	Cobit GAP Interval	Verbal description
9	4.45 - 5.00	Extreme
8	3.89 - 4.44	preference
7	3.34 - 3.88	Very strong
6	2.78 - 3.33	preference
5	2.23 - 2.77	Strong
4	1.66 - 2.22	preference
3	1.12 - 1.65	Moderate
2	0.56 - 1.11	preference
1	0.00 - 0.55	Equal preference

(b) *AHP Score vs. GAP Cobit level*

Alt. A	Alt. B	Cobit GAP	Score
Co. 1	Co. 2	-2	$\frac{1}{4}$
Co. 1	Co. 3	1	2
Co. 1	Co. 4	-3	$\frac{1}{6}$
Co. 1	Co. 5	1	2

(c) *Comparison for Control A.5.1.1*

4.3 Calculation of the Comparison

As mentioned above, the actual calculation of the AHP is done with R. The implementation in R worked with the help of a YAML (Ain't Markup Language) script executed in R. The YAML script is a simplified markup language for data serialization. The YAML script contains all results of the pairwise comparison of criteria and sub criteria, as well as the maturity levels of the 114 controls of the 5 eCommerce companies. The decision hierarchy built up in the YAML script corresponds to the ISO standard. The decision hierarchy is then enriched with alternatives. The paired comparison of the alternatives is executed by a function of the R-package 'ahp' (version 0.2.12 from Christoph Glur) at script runtime for a simple data processing flow. The runtime of the script (with data from 5 companies) on an iMac (3.2 GHz Intel Core i5) was less than 10 s, indicating that it is efficient enough to handle large amounts of data easily.

5 Results of the Comparison

The AHP was used to compare the maturity level in order to find the company with the best information security within an industry (here eCommerce).

Prioritization of Controls. Here we show which priority the control categories (criteria) and controls (sub criteria) have in relation to the complete Appendix A of ISO 27001 over all. The pairwise comparison for the eCommerce industry shows that the controls of the control category 'A.14' have the highest priority (17.6%), followed by 'A.17' (14.7%) and 'A.12' (10.1%). Within control category 'A.14', controls 'A.14.2.8' (22.6%), 'A.14.2.7' (15.2%) and 'A.14.2.6' (11.8%) are the most important as shown in Fig. 2.

	Priority	Company3	Company5	Company1	Company4	Company2
Comparison eCommerce	100.0%					
A.14 System acquisition	17.6%					
A.14.2.8 System security testing	22.6%	16.7%	4.3%	32.7%	32.7%	13.7%
A.14.2.7 Outsourced development	15.2%	23.3%	27.6%	27.6%	17.0%	4.5%
A.14.2.6 Secure development environment	11.8%	19.8%	22.4%	22.4%	13.1%	22.4%
A.14.2.1 Secure development policy	8.0%	4.4%	39.6%	19.4%	19.4%	17.2%
A.14.1.2 Securing application services on public networks	7.4%	5.9%	23.5%	23.5%	23.5%	23.5%
A.14.1.3 Protecting application services transactions	7.0%	6.1%	5.4%	30.3%	30.3%	27.9%
A.14.1.1 Information security requirements analysis and specification	6.2%	21.6%	23.3%	23.3%	23.3%	8.5%
A.14.2.9 System acceptance testing	5.1%	17.5%	19.8%	19.8%	19.8%	23.2%
A.14.2.2 System change control procedures	4.9%	14.3%	28.6%	14.3%	28.6%	14.3%
A.14.2.4 Restrictions on changes to software packages	4.4%	28.6%	14.3%	14.3%	14.3%	28.6%
A.14.2.5 Secure system engineering principles	3.9%	20.0%	20.0%	20.0%	20.0%	20.0%
A.14.2.3 Technical review of applications after operating platform changes	3.5%	28.6%	14.3%	14.3%	14.3%	28.6%
A.17 Information security aspects of business continuity management	14.7%					
A.17.1.2 Implementing information security continuity	48.1%	40.0%	10.0%	20.0%	10.0%	20.0%
A.17.1.1 Planning information security continuity	40.5%	28.6%	14.3%	28.6%	14.3%	14.3%
A.17.1.3 Verify review and evaluate information security continuity	11.4%	40.0%	10.0%	20.0%	10.0%	20.0%
A.12 Operations security	10.1%					
A.12.1.3 Capacity management	13.9%	40.0%	21.9%	21.9%	11.4%	4.8%
A.12.4.1 Event logging	13.2%	35.9%	5.4%	19.6%	19.6%	19.6%
A.12.6.1 Management of technical vulnerabilities	12.4%	19.4%	36.7%	19.4%	19.4%	5.2%
A.12.4.2 Protection of log information	11.5%	24.4%	3.7%	23.7%	23.7%	24.4%
A.12.1.2 Change management	11.3%	20.0%	20.0%	20.0%	20.0%	20.0%
A.12.4.3 Administrator and operator logs	10.2%	5.3%	5.3%	30.5%	30.5%	28.4%
A.12.1.4 Separation of development	9.3%	9.2%	4.5%	29.7%	29.7%	27.0%
A.12.6.2 Restrictions on software installation	7.1%	6.6%	26.3%	12.2%	26.3%	28.7%
A.12.4.4 Clock synchronisation	5.6%	24.3%	12.2%	12.2%	6.7%	44.6%
A.12.1.1 Documented operating procedures	5.6%	5.3%	38.8%	21.7%	21.7%	12.5%

Fig. 2. Top3 control categories prioritized and companies ranked

Comparision of the Companies. The Control Category 'A.14' was used to exemplarily show the evaluation. Figure 2 also shows how the individual eCommerce companies weighting compare with each other in the control category 'A.14' in detail. Overall (cf. Fig. 3), Company3 (21.0%), Company5 (20.9%) and Company1 (20.5%) came out best in a direct comparison. The differences are marginal and only on closer inspection are there more pronounced differences observed at the control level. In relation to a control category e.g. of 'A.14', the maturity of Company1 (4.4%) and Company4 (4.0%) is better in detail, but considering the control category 'A.17', Company3 (5.2%) is clearly ahead of Company4 (1.7%).

	Weight	Company3	Company5	Company1	Company4	Company2
Comparison eCommerce	95.1%	21.0%	20.9%	20.5%	16.3%	16.2%
A.14 System acquisition	17.6%	3.0%	3.3%	4.4%	4.0%	3.0%

Fig. 3. Control category A.14 weight contribution and ranked companies

6 Discussion

Based on these results, we discuss the main findings as follows. The results show that with the pairwise comparison it is possible to obtain a priority for each individual control, and thus very granular, in the overall context of ISO/IEC 27001 for the eCommerce industry. The priorities of the larger control categories are also very helpful, as a quick comparison of priorities is possible here. The approach with the pairwise comparison by AHP meets all requirements of the methodology part. Similarly, it is shown that the weighting of the pairwise comparisons of the maturity level of eCommerce companies can be mapped very granularly to the controls of the ISO/IEC 27001 standard. It was also possible to derive the AHP score from the maturity levels automatically. This makes it easy to compare the rankings of the companies. The only effort which needs to be invested (for each industry) is the prioritization of the controls.

The results suggest that the approach works in conjunction with real data (the maturity levels of HBM's eCommerce companies) at least for the chosen area. The results of the comparison also withstand the reality that one of the authors observes in his daily professional life. The results also showed that the ranking results reflect the reality of at least the HBM eCommerce companies. However, it can be strongly assumed that the method is directly applicable to other companies with the same or similar results.

6.1 Limitations

For reasons of simplification and clarity, we have demonstrated the approach only with a small number of companies. But is easily possible to run the approach with the full set of HBM's companies and to extend it to other business units by readjusting the ISO/IEC 27001 controls' priorities.

The application of the AHP methodology is not undisputed in technical literature. At this point the authors consider some points of this criticism. On the one hand, these are points concerning the mathematical part of the AHP and on the other hand, the criticism is based on the procedure. In the model calculated above, the pairwise comparison of the criteria and sub criteria has been carried out by one person (with expert knowledge), which can be regarded as a very subjective survey of all pair comparisons. This assumes that there are high demands on the respondent due to the many pair comparisons, which is why there are often problems with validity [18]. This could lead to a limitation of the size of the decision model and is seen as a critical and possible optimization point of the AHP methodology in literature and practice [11].

If you take a closer look at the origin of the maturity level, you immediately notice that it is determined by the information security officer's self-disclosure. As with all quantification, the human factor, a lack ob objectivity or bias, cannot be excluded here. However, it can be largely validated by a team of experts. Another point concerns the type of data collection, the resulting prevailing data quality and possible imponderables in data evaluation. These issues could only be reduced but not completely eliminated by several iterations of quality assurance.

In the next chapter, some of the limitations will be discussed and further improvements of the methodology/model will be proposed.

7 Conclusion and Future Work

The results of the pairwise comparison suggest that AHP is very well suited to compare the information security maturity of different companies and to find the company with the best information security within an industry.

It has been proven that a comparison within the eCommerce industry is possible using this model and thus ranking the prioritization of control categories and, above all, the individual controls can follow. The AHP provides in this case a robust and comprehensive treatment for decision makers in both qualitative and quantitative ways as found in this study and it can be assumed that this will also work for other companies in the same environment. The real insight is to adapt the AHP or the data so that it works together. The AHP-model has shown how AHP might be used to assist decision maker evaluate information security in one branch. Very interesting, and also for validation, would be the pairwise comparison for other industries such as publishing houses, manufacturing industry. Companies with very different degrees of maturity could also be interesting here.

Some of the limitations mentioned above regarding the AHP methodology deal with the comparison of pairs. A possible improvement of the model would be to compare it with the help of a team of experts from the eCommerce industry. This would have the advantage that the pair comparison is subject to validation.

In future work, the focus will be on the details of implementing this model across a variety of different examples, as well as working on more expanded decision hierarchy with an additional level of sub criteria (control objectives). In addition, it would be interesting to calculate the approach with different aggregated data (min, max, median) in addition to the mean value and to observe the effects. Furthermore, it would be interesting to apply the AHP methodology to other industries (e. g. publishing, manufacturing industry etc.). Ultimately, this would provide the prerequisites for comparing information security across industries, comparing apples and pears, so to speak.

References

1. Abbas Ahmed, R.K.: Security metrics and the risks: an overview. Int. J. Comput. Trends Technol. **41**(2), 106–112 (2016)
2. Al-Shameri, A.A.N.: Hierarchical multilevel information security gap analysis models based on ISO 27001: 2013. Int. J. Sci. Res. Multidisc. Stud. **3**(11), 14–23 (2017)
3. Anderson, R., et al.: Measuring the cost of cybercrime. In: Böhme, R. (ed.) The Economics of Information Security and Privacy, pp. 265–300. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-39498-0_12
4. Axelrod, C.W.: Accounting for value and uncertainty in security metrics. Inf. Syst. Control J. **6**, 1–6 (2008)

5. Bodin, L.D., Gordon, L.A., Loeb, M.P.: Evaluating information security investments using the analytic hierarchy process. Commun. ACM **48**(2), 78–83 (2005)
6. Böhme, R.: Security metrics and security investment models. In: Echizen, I., Kunihiro, N., Sasaki, R. (eds.) IWSEC 2010. LNCS, vol. 6434, pp. 10–24. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-16825-3_2
7. Choo, K.K., Mubarak, S., Mani, D., et al.: Selection of information security controls based on AHP and GRA. In: Pacific Asia Conference on Information Systems, vol. 1, no. Mcdm, pp. 1–12 (2014)
8. Eisenführ, F., Weber, M.: Rationales Entscheiden, p. 415. Springer, Heidelberg (2003). https://doi.org/10.1007/978-3-662-09668-0
9. Gordon, L.A., Loeb, M.P.: The economics of information security investment. ACM Trans. Inf. Syst. Secur. **5**(4), 438–457 (2002)
10. Haufe, K.: Maturity based approach for ISMS. Ph.D. thesis, University Madrid (2017)
11. Ishizaka, A., Labib, A.: Review of the main developments in the analytic hierarchy process. Expert Syst. Appl. **38**(11), 14336–14345 (2011)
12. ISO/IEC 27001: Information Technology—Security Techniques—Information Security Management Systems—Requirements. International Organization for Standardization (2013)
13. Khajouei, H., Kazemi, M., Moosavirad, S.H.: Ranking information security controls by using fuzzy analytic hierarchy process. Inf. Syst. e-Bus. Manag. **15**(1), 1–19 (2017)
14. Le, N.T., Hoang, D.B.: Capability maturity model and metrics framework for cyber cloud security. Scalable Comput.: Pract. Exp. **18**(4), 277–290 (2017)
15. Lee, M.C.: Information security risk analysis methods and research trends: AHP and fuzzy comprehensive method. Int. J. Comput. Sci. Inf. Technol. (IJCSIT) **6**(February), 29–45 (2014)
16. Liu, D.L., Yang, S.S.: An information system security risk assessment model based on fuzzy analytic hierarchy process. In: 2009 International Conference on E-Business and Information System Security, pp. 1–4 (2009)
17. Majumder, M.: Impact of Urbanization on Water Shortage in Face of Climatic Aberrations. Springer, Singapore (2015). https://doi.org/10.1007/978-981-4560-73-3
18. Millet, I.: Ethical decision making using the analytic hierarchy process. J. Bus. Ethics **17**(11), 1197–1204 (1998)
19. Mu, E., Pereyra-Rojas, M.: Pratical Decision Making: An Introduction to the Analytic Hierarchy Process (AHP) Using Super Decisions (v2). Springer, Heidelberg (2017). https://doi.org/10.1007/978-3-319-33861-3
20. Nasser, A.A.: Measuring the information security maturity of enterprises under uncertainty using fuzzy AHP. Int. J. Inf. Technol. Comput. Sci. **4**(April), 10–25 (2018)
21. Peters, M.L., Zelewski, S.: Analytical Hierarchy Process (AHP) – dargestellt am Beispiel der Auswahl von Projektmanagement-Software zum Multiprojektmanagement. Institut für Produktion und Industrielles Informationsmanagement (2002)
22. Rudolph, M., Schwarz, R.: Security indicators – a state of the art survey public report. FhG IESE VII(043) (2012)
23. Saaty, T.L., Vargas, L.G.: Decision Making with the Analytic Network Process: Economic, Political, Social and Technological Applications with Benefits, Opportunities, Costs and Risks. Springer, Heidelberg (2006). https://doi.org/10.1007/0-387-33987-6

24. Saaty, T.L., Vargas, L.G.: Models, Methods, Concepts & Applications of the Analytic Hierarchy Process, vol. 175. Springer, Heidelberg (2012). https://doi.org/10.1007/978-1-4614-3597-6
25. Syamsuddin, I., Hwang, J.: The application of AHP to evaluate information security policy decision making. Int. J. Simul.: Syst. Sci. Technol. **10**(4), 46–50 (2009)
26. Vaughn, R.B., Henning, R., Siraj, A.: Information assurance measures and metrics - state of practice and proposed taxonomy. In: Proceedings of the 36th Annual Hawaii International Conference on System Sciences, HICSS 2003 (2003)
27. Watkins, L.: Cyber maturity as measured by scientific-based risk metrics. J. Inf. Walfare **14.3**(November), 60–69 (2015)

B.5 Aggregating Corporate Information Security Maturity Levels of Different Assets

Michael Schmid and Sebastian Pape. Aggregating corporate information security maturity levels of different assets. In *Privacy and Identity Management. Data for Better Living: AI and Privacy - 14th IFIP WG 9.2, 9.6/11.7, 11.6/SIG 9.2.2 International Summer School, Windisch, Switzerland, August 19-23, 2019, Revised Selected Papers*, number 576 in IFIP Advances in Information and Communication Technology, pages 376–392. Springer Boston, 2019. doi: 10.1007/978-3-030-42504-3_24. URL https://link.springer.com/chapter/10.1007/978-3-030-42504-3_24

Aggregating Corporate Information Security Maturity Levels of Different Assets

Michael Schmid[1,2(✉)] and Sebastian Pape[1(✉)]

[1] Chair of Mobile Business & Multilateral Security, Goethe University Frankfurt, Frankfurt, Germany
{michael.schmid,sebastian.pape}@m-chair.de
[2] Hubert Burda Media Holding KG, Munich, Germany

Abstract. General Data Protection Regulation (GDPR) has not only a great influence on data protection but also on the area of information security especially with regard to Article 32. This article emphasizes the importance of having a process to regularly test, assess and evaluate the security. The measuring of information security however, involves overcoming many obstacles. The quality of information security can only be measured indirectly using metrics and Key Performance Indicators (KPIs), as no gold standard exist. Many studies are concerned with using metrics to get as close as possible to the status of information security but only a few focus on the comparison of information security metrics. This paper deals with aggregation types of corporate information security maturity levels from different assets in order to find out how the different aggregation functions effect the results and which conclusions can be drawn from them. The required model has already been developed by the authors and tested for applicability by means of case studies. In order to investigate the significance of the ranking from the comparison of the aggregation in more detail, this paper will try to work out in which way a maturity control should be aggregated in order to serve the company best in improving its security. This result will be helpful for all companies aiming to regularly assess and improve their security as requested by the GDPR. To verify the significance of the results with different sets, real information security data from a large international media and technology company has been used.

Keywords: Information security · Information security management · ISO 27001 · Aggregation functions · Information security controls · Capability maturity model · Security maturity model · Security metrics framework

1 Introduction

Approximately 18 months ago the General Data Protection Regulation (GDPR) containing requirements regarding the processing of personal data of individuals

Published by Springer Nature Switzerland AG 2020
M. Friedewald et al. (Eds.): Privacy and Identity 2019, IFIP AICT 576, pp. 376–392, 2020.
https://doi.org/10.1007/978-3-030-42504-3_24

became operative. The GDPR states that organizations must adopt appropriate policies, procedures and processes to protect the personal data they hold. Article 32 of the GDPR specifically requires organizations to ensure confidentiality, integrity, availability and resilience (core principles of the information security) of processing systems and services, and to implement a process for regularly testing, assessing and evaluating the effectiveness (e.g. with KPIs) of technical and organizational measures for ensuring secure processing [27]. Thus, in addition to presenting a state of the art security level, this article emphasizes the importance of a process for regularly testing, assessing and evaluating the security. However, it does not provide detailed guidance on how to achieve these goals.

It is difficult to judge whether the security level is sufficient from a management perspective. Managers often act according to the maxim 'minimal effort maximum success', since the budget is usually limited. Of course, this also applies to the area of information security and varies depending on the industry and the self-perception of IT security within it. This is justifiable from an economic point of view, but it has an influence on how information security is dealt with in the company. In this situation, it is important to create transparency regarding the state of information security, within an organization to determine how good the process is, as well as in comparison to other companies operating in the same environment. This transparency can be used to demonstrate/ensure that (information) security does not suffer from budget constraints.

An established way to monitor and steer the information security is the implementation of an information security management system (ISMS). With the most popular standard in this field, ISO/IEC 27001 [14], it is possible to manage the information security in a company through the ISO-controls. An effective ISMS that conforms to ISO/IEC 27001 meets all requirements of GDPR's article 32.

The information security status of an environment like a company is a very individual observation [1]. To estimate the actual status of information security normally metrics or key performance indicators (KPI) are taken into account [21]. The information gathering of these KPIs is usually done through different technical or organizational metrics of a company. Using KPI/Metric/-Maturity for the status of information security is only an indicator of improvement or deterioration since there is unfortunately no gold standard for this [4]. It would be very complex and expensive to first collect or generate these KPIs for this evaluation. It is important therefore, to work with the data/metrics already available and no need for further data collection. In this context, it should not go unmentioned that another standard exists in this environment, the ISO/IEC 27701 [15]. This standard deals with how to establish and run a Privacy Information Management System (PIMS) that adds Personally Identifiable Information (PII) security protection to an existing ISMS. In order to assess the status of information security as well as the quality of the process, mostly a maturity model is used. A common method for the assessment of the maturity is the COBIT control maturity model from the ISACA framework [13]. With the help of this model it is possible to assess the goodness of the ISO-controls on a 0 to 5 scale. The assessment supports the improvement of the organization's security and delivers the management perspective in the fulfillment of regulatory requirements.

With the maturity level, the manager has a relatively good overall view of the status of information security. However, this is usually a very aggregated view of the status, as a company will operate different types of IT systems/applications to support its business process. The information assets worth protecting (e.g. customer data, trade secrets, source code, etc.) are not only processed or stored on one IT system, but on several. As a consequence, the maturity level may differ between systems. Therefore, many companies not only collect a maturity level for the whole company, but also a maturity level per system for each control [11]. An ISO control such as A.12.6.1 (Vulnerability Management) will only be able to reflect a combined value from several IT systems/applications. That's why, different values exist for different assets per ISO control (see Fig. 1).

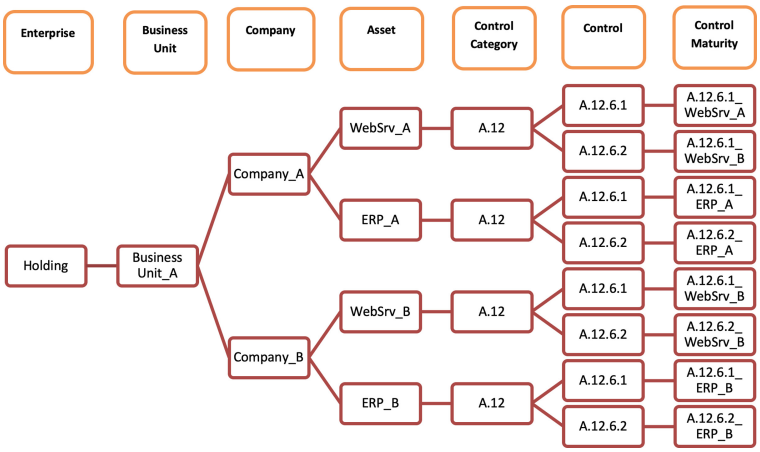


Fig. 1. Exemplary holding structure with different assets and control maturity for ISO-controls

In order to derive a KPI from the assets' control maturity level or use them as input for existing approaches [24,25], the questions arises how they can be meaningfully aggregated.

RQ1: How can maturity levels for one control be meaningfully aggregated across different assets?
Different aggregation types can not only influence the outcome of the approach, but also influence the managers which security controls should be improved.
RQ2: How would a manager's optimization strategy depend on the different aggregation methods?
And finally, it's equally important to consider the aggregation's influence on the final result of the algorithm.
RQ3: How much does the outcome of a holistic approach actually change depending on different aggregation types?

To examine this research question, we first discuss different types of aggregation for maturity levels. In the next step, for each of the aggregations we derive possible security managers' optimization strategies in order to establish which control to improve next. For a reality check, we examine asset's maturity levels from real company data to check if our assumptions are realistic. As a final step, we also use real companies' maturity levels to examine how much the outcome of [24] would be changed by applying a different aggregation.

The remainder of this work is structured as follows: In Sect. 2 we give a brief overview of related work. Section 3 describes our methodology how we developed our approach for each research question shown in Sect. 4. Our results are shown in Sect. 5 and discussed in Sect. 6, respectively Sect. 7.

2 Background and Related Work

In addition to the differences in the assessment of information security, all assessment procedures also have in common that the ratings of the maturity level and the weighting of weights are not allocated to a common overall value in the sense of an 'information security score'. It is, therefore up to the evaluator to carry out the respective evaluation, as he or she is forced to choose between these two quantitative aspects of the evaluation, e.g. the ratings on the one hand and the weighting on the other [17]. Savola [23] discussed a broader approach to finding a metrics which can be used in the field of different security disciplines like management and engineering practices. In contrast to this, the works of Böhme [8] and Anderson [4] deal more with the economic impact of investments in information security. There are also other models that deal with the measurement of information security using maturity levels e.g. the Information security maturity model (ISMM) [22] and the Open Information Security Maturity Model (O-ISM3) [22]. ISMM is intended as a tool to evaluate the ability of organizations to meet the objectives of security and O-ISM3 aims to ensure that security processes operate at a level consistent with business requirements. However, both models refer more to the process level than to the asset level. The focus of this work is to compare the different aggregation types of maturity within an industry. This could later lead to a monetary assessment of information security or maturity.

2.1 Aggregation Types

Unfortunately, the precise process of how to aggregate maturity levels is neither well documented nor comprehensively studied or understood (from a psychological perspective), so most of this labor is done by rule-of-thumb [26]. As mentioned, our approach varies between four aggregation types - namely the minimum, maximum, average and median - to compare their different potential impacts on decision making. Regarding the two measures of central tendency (average, median), strengths and weaknesses have been discussed in scientific literature. Averages are strongly influenced by extreme values. In our context, this could lead to an over- or underestimate of control maturity. In contrast,

the median is not skewed by extreme values, consequently running the risk of overestimating control maturity [10]. The opposite can be the case when there are multiple non-values (e.g. zeros) in a data sample, as laid out by Anderson et al. [5]. The relative position of average and median differs in skewed distributions. A distribution skewed to the left will lead to a smaller median compared to the average, while a right-skewed distribution reverses the relation [18]. Overall, it makes sense to include both measures of central tendencies in our analysis to compensate for weaknesses and bias. The minimum and maximum further alleviate potential misrepresentations of control maturity, as they provide the numerical range of scores and expose potential outliers [7]. Logically, both measures are most sensitive to outliers in a data set but are nevertheless useful in our analysis when used in combination with the measures of central tendency.

2.2 Aggregation of Security Metrics

Although the domain of security metrics has been covered by a number of authors [3], only limited work on the area of metrics aggregation has been carried out. Ramos et al. [20] provided a detailed survey on models for quantifying networks resilience to attacks. The authors used stochastic techniques and attack graphs to map the possible routes an attacker could take to compromise a system. Abraham et al. [2] discussed the challenges faced by practitioners in the field of security measurements and highlighted the need to develop a mechanism for quantifying the overall security of all the systems on the network. The authors proposed a predictive framework that uses stochastic techniques based on attack graphs and incorporated temporal factors relating to the vulnerabilities such as availability of patch and exploits predicting the future state of the system. Cheng et al. [9] proposed a model for aggregating security metrics using Common Vulnerability Scoring System (CVSS) base metrics to estimate the exploitability of the vulnerabilities. Homer et al. [12] and Beck et al. [6] proposed a mathematical security model for aggregating vulnerabilities in risks in enterprise networks based on attack graphs. An aggregated numeric value was assigned to show the likelihood of a vulnerability being exploited by an attacker.

3 Research Methodology

The general aim of our approach is to determine which effect the different aggregation types of the maturity control of assets have on the information security of the companies. In order to do this it is important to create transparency around the state of information security. The method should take into account the different requirements of the different research questions set out in Sect. 1.

We derive the different aggregation methods in the next subsection for our approach, then determine the proper algorithm and finally describe the data collection of our approach.

3.1 Different Aggregation Functions

First, we examine which functions are suitable to verify the approach described above. As shown in Sects. 2.1 and 2.2 with the different aggregation functions e.g. average, median, minimum and maximum it is possible to form a single summary value from a group of data. The challenge now is to find the right aggregation functions to support the approach provided. These aggregation functions have in common that they can represent the impact of decisions by information security managers, each type in its own way. The hypotheses provide an outlook how information security managers might behave in terms of aggregation.

3.2 Data Collection

It would be very complex and expensive to first collect or generate these KPIs for this evaluation. It is important to use data/metrics already available (e.g. information security maturity level). To test the above approach it is necessary to set up the model and verify it with real data. We need a maturity assessment of the ISO/IEC controls and to weight and aggregate them according to the specific industry. We focused on the eCommerce industry for the following reasons:

- Available data from a large range of companies
- Excellent data quality and validity
- High actuality of the existing data
- Very good know-how available in the expert assessment of the industry

We collected data from Hubert Burda Media (HBM), an international media and technology company (over 12,500 employees, more than 2.5 billion annual sales, represented in over 20 countries). This group is divided into several business units that serve various business areas (including print magazines, online portals, eCommerce etc.). The business units consist of over 250 individual companies with about 30 of them being in the eCommerce industry. Each subsidiary operates independently of the parent corporation. There is a profit center structure, so the group acts as a company for entrepreneurs and the managing directors have the freedom to invest money in information security and to choose the appropriate level of security. We will briefly describe how this data is collected before going into more detail on the data used for the comparison. Each individual company in the group operates its own Information Security Management System (ISMS) in accordance with ISO/IEC 27001, which is managed by an Information Security Officer (ISO) on site and managed by a central unit in the holding company. As part of the evaluation of the ISMS, the maturity level of the respective ISO 27001 controls is ascertained - very granularly at the asset level (application, web-server, CRM etc.). The maturity level is collected/updated regularly once a year as part of a follow-up procedure.

3.3 Algorithm Method Selection

Taking all requirements of the method into account, a previously developed approach from Schmid and Pape [24] is applicable. The primary objective of this

approach was to show how to use the analytic hierarchy process (AHP) to compare the information security controls of a level of maturity within an industry in order to rank different companies. The AHP is one of the most commonly used Multiple Criteria Decision Methods (MCDM), combining subjective and personal preferences in the information security assessment process [19]. It allows a structured comparison of the information security maturity level of companies with respect to an industry [26] and to obtain a ranking [16]. This allows the definition of a separate weighting of information security metrics for each industry with respect to their specifics while using a standardized approach based on the maturity levels of the ISO/IEC 27001 controls.

To achieve the aim of this paper it is necessary to calculate the control maturity of the assets with different aggregation types such as: minimum, maximum, average or median. This shows how strong the characteristics of the individual aggregation types are in comparison to the real data. Out of this, the first indicators can then be derived to clarify which effect the aggregation types have on the information security for individual companies. The following chapter describes the implementation of the approach for each of the 3 research questions.

4 Discussion of Different Aggregations

As outlined in the previous chapter the different aggregation functions have a very likely a different outcome when it comes down comparing them with each other. Among other things, this chapter will describe the different characteristics of the aggregation functions as well as the effects of the various IT assets of a company and how they affect the results. A vivid example with real world data illustrates how the various aggregations affect the final result and ultimately the behaviour of those responsible for information security.

4.1 General Aggregation Functions

The great advantage of the aggregation functions average, median, minimum and maximum is that by aggregating (key) figures differences can be identified in the results and thus comparisons can be made. These could be a strength or weakness per each aggregation type. In contrast to this, there is no difference in the comparison of the results for the aggregation functions sum, range and count, for example. A further advantage of the four aggregation functions mentioned above is the adaptability of these types to a different number of values. They work nicely even if each company has a different number of assets considered. This makes it possible to derive different scenarios for the comparison.

4.2 Derived Optimization Strategies

If the results of the different aggregation functions are compared with each other, different optimization strategies can be derived in the end. This is particularly

Table 1. Maturity levels of different collective assets for the ISO-control A.12.6.1 from five companies

Asset	Company1	Company2	Company3	Company4	Company5
1	4	**0**	3	3	4
2	4		**2**	2	**4**
3	**4**		2	3	
4	1			1	
5	0				

Table 2. Maturity level results from different aggregation functions

Aggregation	Company1	Company2	Company3	Company4	Company5
average	2.6	0	2.3	2.25	4
median	4	0	2	2.5	4
minimum	0	0	2	1	4
maximum	4	0	3	3	4

important for those who are responsible for information security. Due to the different aggregations, it is possible that different optimization possibilities can be shown in the evaluation of information security. The information security manager can then decide which optimization strategy/aggregation function brings him the most benefit. If we take a closer look at the 4 aggregation functions mentioned above and examine them for the possible outcome, we obtain the following hypotheses:

- minimum → improve only the worst value (weakest chain, can make sense),
- maximum → improve only the best value (is this desirable?),
- average → improve any value (probably the easiest ones first) and
- median → may lead to a really two-fold security level with $\frac{n-1}{2}$ insecure services and $\frac{n+1}{2}$ secure services.

As next step we validate these hypotheses using an example with real world data.

4.3 Example with Real World Data

In order to compare the results of the different aggregation functions we need real data. Section 3.2 describes how these real data, in this case the COBIT maturity, are collected. For a concrete example we use the maturity level for a specific ISO-Control (here A.12.6.1 'Management of Technical Vulnerabilities') because this control focuses on an IT asset. As an example, we use data from five companies and their various IT assets (see Table 1).

Based on this data, the calculations of the four different aggregation functions are now performed (see Table 2) for the five companies. The colored cells highlight the aggregation functions and the maturity levels used. These exemplary calculations are based on the maturity levels of companies with different IT assets. A company uses many different IT assets to support its core and support processes. The next chapter examines these different types of IT assets in more detail.

4.4 More Complex Aggregations

In order to steer manager's optimization strategy one needs to integrate weightings for the different assets. This leads to the problem that many approaches, e.g. AHP [24] only work with a fixed number of assets. Considering only a fixed set of assets for each domain would narrow the defined scope, thus it should be possible to still evaluate a different number of assets. Conclusion: Define most important assets and their weighting and build an asset class for all remaining assets. This way, at least the impact of the manager's optimization strategies is more limited and only usable among the assets within the 'special class'. Arising Question: How to derive the priorities for all the classes?

When considering the core business processes for an eCommerce company, the web presence, a merchandise management system and a customer management system are normally expected. For this stage, we examined the prevailing situation of the IT assets used by 25 eCommerce companies from HBM and evaluated them. Almost all eCommerce companies had a web sever (24), a database server (24), an ERP system (22) and a CRM system (20). Further IT assets, which did not have such a high frequency were mail servers (14), file servers (14), dev servers (12), git (9), ftp servers (7), etc. This also coincides with the assumption resulting from the core business processes. Resulting from this the core IT assets of an eCommerce company, a web sever, a database server, an ERP system and a CRM system were selected.

Only considering these core IT assets would not reflect the overall picture of an eCommerce company. In order to have a comprehensive picture we also need the assets that are used in the IT department (e.g. file server, dev server, ftp server etc.). We have combined these IT assets into one collective asset for the comprehensive picture. In a further step, this collective asset, or better the maturity level, is calculated or evaluated using various aggregation types (minimum, maximum, average, median). In combination with the 4 core assets, aggregated values of the collective assets are included in the calculation as 5th assets (with 20%). This can provide the first insights as to whether a certain aggregation method might influence the units or sub-companies decision, hence which control should be improved next.

4.5 Priorization of Asset Classes

The core IT assets are equally important (e.g. 25% for each) at the moment. An interesting question would be e.g. how much more important is the web server

of an eCommerce company compared to the ERP system? It would be necessary to add an additional layer of prioritization in order to differentiate between the differing control requirements. In order to implement this we could use the CIA triad model which encompasses a triangle of tension between the three principles Confidentiality, Integrity and Availability. When applied to our use case, the principles of importance vary between control objectives and is represented by a score for the CIA principles according to their importance for these control objectives. This would provide for an extension of the approach by the CIA values of the individual assets. In order to do this, we need the CIA evaluation per IT asset. The information (e.g. customer data, contracts etc.) is stored or processed on an IT asset. It allows conclusions to be drawn as to how this asset should be treated in terms of confidentiality, integrity and availability. This means that there is at least one information asset per asset, but usually several information assets per asset, which are evaluated according to the CIA criteria with a 3-step classification (normal, advanced and high). A web server will, for example, process or even store information assets such as customer data, bank details, etc. If the information values 'customer data' and 'bank details' for a web server are uniformly evaluated for confidentiality, integrity and availability according to a given system, this can be set in relation to an ERP system with the information values 'purchasing conditions' and 'master data'. A further step was needed to convert our CIA data to pairwise comparisons on our AHP score, as depicted in Table 3a. We define a factor of equal importance regarding the CIA triad of all four core assets as a proportion percentage of 25% each. Consequently, we can conduct pairwise comparisons related to the proportion gaps in our data, which are then normalized based on the AHP preference score i.e. equal importance (AHP score: 1) is expressed by tiny differences in proportion to percentage of smaller than 2.77%, while the highest order of relative importance (AHP score: 9) means a difference of 25% in proportion to percentage (see Table 3b).

Table 3. Combined GAP of core assets and AHP Score

AHP Score	Verbal description
9	Extreme preference
8	
7	Very strong preference
6	
5	Strong preference
4	
3	Moderate preference
2	
1	Equal preference

(a) Fundamental AHP Score

AHP Score	Proportional CIA differences	Verbal description
9	22.22 - 25.00	Extreme preference
8	19.45 - 22.21	
7	16.67 - 19.44	Very strong preference
6	13.89 - 16.66	
5	11.12 - 13.88	Strong preference
4	08.34 - 11.11	
3	05.56 - 08.33	Moderate preference
2	02.78 - 05.55	
1	00.00 - 02.77	Equal preference

(b) AHP Score vs. GAP of the CIA differences

5 Results of the Holistic Approach Considering Different Aggregation Types

The aim of this paper is to find out which effects the different aggregation functions have on the results and which conclusions can be drawn from them. The different aggregation functions can not only influence the outcome of the approach, but also influence the manager's decision as to the order in which control's maturity levels should be increased. They can influence the manager's optimization strategy depending on the different aggregation functions. At present, the maturity levels have not yet been examined with a view to optimization.

Table 4. Comparison of different aggregation types from 5 companies only for control A.12.6.1

Aggregation/proportion	Company1	Company2	Company3	Company4	Company5
Average	15.4%	7.7%	30.8%	30.8%	15.4%
Median	12.6%	12.6%	27.4%	34.9%	32.0%
Minimum	10.0%	10.0%	40.0%	20.0%	20.0%
Maximum	22.2%	11.1%	22.2%	22.2%	22.2%

Table 5. Comparison (proportion) of different aggregation types from 5 companies for control category A.12

Aggregation	Company1	Company2	Company3	Company4	Company5
Average	1.7% *(17.9%)*	1.2% *(12.6%)*	2.3% *(24.2%)*	2.1% *(22.1%)*	2.2% *(23.1%)*
Median	1.6% *(16.8%)*	1.7% *(17.9%)*	2.4% *(25.3%)*	1.9% *(20.0%)*	1.9% *(20.0%)*
Minimum	1.4% *(14.7%)*	1.2% *(12.6%)*	2.8% *(29.5%)*	2.1% *(22.1%)*	2.0% *(21.0%)*
Maximum	1.8% *(18.9%)*	1.3% *(13.7%)*	1.7% *(17.9%)*	1.6% *(16.8%)*	3.1% *(32.6%)*

5.1 Results of Aggregated Maturity Levels

The AHP was used to compare the maturity levels in order to work out how a maturity control should be determined to best serve the company in improving its security with reference to the first research question [24]. Table 4 shows a comparison of results with different aggregation types from five companies only for control A.12.6.1 'Management of Technical Vulnerabilities'. Because this control is asset-based, this value is composed of different IT assets that were calculated with each of the 4 different aggregation types.

As expected, Company 2 is very weakly developed if the raw data in Table 1 is considered. Company 1 is also quite clearly recognizable with regard to the minimum and maximum. Company 3 has the highest proportion concerning the

minimum (40.0%). The results show that a detailed look at Company 5 would be worthwhile, as the largest fluctuations between average and median (15.4%–32.0%) can be observed here.

If we now abstract this comparison to a higher level, e.g. no longer to the control level but to control category level, the results should no longer fluctuate greatly. In the case of control categories, we are concentrating only on the most important ones for the eCommerce industry. The weighting of the respective control categories can be seen from the results of the AHP [24]. 'A.14' (System Acquisition, Development and Maintenance) is the most important for the eCommerce industry with 16.5%, followed by 'A.17' (Information Security Aspects of Business Continuity Management) with 14.7% and then 'A.12' (Operations security) with 9.5%. Table 5 shows how the individual eCommerce companies weighting is compared with each other and the four different aggregation types for 'A.12' Operations security are compared in detail.

Table 6. Comparison of different aggregation types from 5 companies for the complete ISO/IEC 27001

Aggregation/proportion	Company1	Company2	Company3	Company4	Company5
Average	16.7% *(4.)*	15.4% *(5.)*	19.8% *(1.)*	18.3% *(3.)*	19.5% *(2.)*
Median	16.7% *(4.)*	16.3% *(5.)*	19.8% *(1.)*	18.8% *(2.)*	18.1% *(3.)*
Minimum	16.6% *(4.)*	14.6% *(5.)*	21.3% *(1.)*	18.7% *(2.)*	18.5% *(3.)*
Maximum	17.5% *(2.)*	15.6% *(5.)*	16.1% *(4.)*	16.2% *(3.)*	24.2% *(1.)*

The rows total up to 9.5% because it is the ratio of 'A.12' weighting in contrast to the overall control categories. The distribution of values within an aggregation type per company is specified in brackets. The differences are marginal but a closer inspection more pronounced differences can be observed at the control level and therefore tendencies are recognizable. Company 3 has again the highest proportion concerning the minimum (29.5%)

The last comparison in this environment is the application of the four different aggregation types to the complete controls of Annex A of ISO/IEC 27001. This is ultimately the highest expected level of aggregation of this approach. It is to be expected that the results will no longer differ so much from each other. Table 6 shows the results of the comparison.

The rows total up only to 89.9% because 11.1% is a 'measure of the error due to inconsistency' which is provided by the AHP. The ranking within all companies is specified in brackets. Concerning the outcome of the comparison, Company 5 stands out with a high value for maximum aggregation (24.2%) and Company 1 looks very stable concerning the different aggregation types. Generally, the minimum does not fluctuate as much as the maximum. Company 1 to 3 have no high fluctuation in common and concerning Company 3 there is not a lot of variance can be observed.

5.2 Results of Priorization the Asset

The descriptive statistic of HBMs information asset presence is used to begin with the set of four core assets, namely web server (24), database server (24), ERP system (22) and CRM system (20). Besides, computing our input scores as well as defining our priorities for sub criteria level requires the processing of the CIA inputs. The summarizing statistic is presented in Table 7 below.

All CIA scores are summed up for each asset and divided by the total number (see Table 8). The lowest sum resulted from the CRM asset with 100, and is hence our base value.

Concerning the priorization of asset classes Table 9 shows a pairwise comparison of the core assets from one eCommerce company. The deviation is then transformed into the AHP scores with the help of the intervals from the GAP of core assets (see Table 3b). It is clear that the biggest difference lies between the web server and the CRM system (11.7%) and the smallest difference between

Table 7. CIA of information assets from different IT assets of one company

Company	Information asset for	Confidentiality	Integrity	Availability	Sum of CIA
Company_1	Web-Server	2	2	3	**7**
	Web server	3	3	3	**9**
	Web server	3	3	2	**8**
	Web server	2	3	2	**7**
	Web server	3	3	2	**8**
	Database server	2	2	2	**6**
	Database server	2	2	2	**6**
	ERP system	2	2	2	**6**
	ERP system	2	2	2	**6**
	ERP system	2	2	2	**6**
	ERP system	2	2	2	**6**
	CRM system	2	2	2	**6**
	CRM system	2	2	2	**6**
	CRM system	1	2	2	**5**
	CRM system	1	2	2	**5**
Company_2	...	...	...	...	...

Table 8. Distribution of assets

Asset	CIA sum	Distribution
WEB	156	32.5%
ERP	104	25.0%
DB	120	21.7%
CRM	100	20.8%

the CRM system and the database server (0.7%). With the help of this score it is possible to weight the core assets based on their CIA assessment and process them with the AHP.

6 Discussion

Based on these results, we discuss the main findings as follows. The results show that it is possible to elaborate differences in the assessment and comparison of IT assets with the help of different aggregation types. The main goal of this paper, to assist managers in how they can improve their information security by comparing different aggregated information security maturity levels on asset level has shown several outcomes. The results show that a certain type of aggregation affects a company when trying to improve its maturity levels (see Table 4). Company 1 and 2 would improve first the collective assets with a low control maturity if a minimum aggregation is used. If the aggregation function maximum is used Company 3 would try to improve one collective asset in order to maximize only one control maturity (see Table 5). Concerning the big picture in Table 6 the ranking of the companies differs only for Company 1 and 3. Company 1 has already very high control maturities, so it is not as easy for them to improve. Company 3 almost a very homogenous control maturity thats why the would probably improve only one collective assets if the maximum aggregation is chosen. The other companies are more or less stable concerning the ranking, e.g. Company 2 does not changes at all.

Table 9. AHP Comparison with core assets

Sub criteria A	Sub criteria B	A/B	Deviation	Score
WEB	ERP	A	+7.25%	3
WEB	DB	A	+10.8%	4
WEB	CRM	A	+11.7%	5
ERP	DB	A	+2.3%	1
ERP	CRM	A	+4.1%	2
DB	CRM	A	+0.7%	1

With the help of the CIA prioritization is possible to first weight and then aggregated the different IT systems and applications with each other (see Table 9). The results show hat for an eCommerce company it is obvious that the web server is more important than the ERP-System in supporting the business processes.

6.1 Limitations

Maturity levels are not assessed automatically but by each of the individual companies' information security officer (ISO). Therefore, there may be discrepancies in the way the maturity levels are understood and assessed. This is clearly a limitation of any approach based on security maturity levels, but it might limit the informative value of the collected maturity levels. Moreover, the maturity levels are reported to the management and they result in a key performance indicator (KPI) for security for that specific unit. Thus, it can be assumed that each ISO has an interest in having a good evaluation. Therefore, ISOs might be tempted to assess the maturity levels more optimistically or to limit the scope of the information security management system in order to achieve better evaluations more easily. A common understanding of the different maturity levels is already established by guidelines and manuals provided to the ISOs (of HBM). This could be expanded further in order to reach a better understanding for the assessment of control maturity levels. Furthermore, deviations can be addressed if the companies are (externally) audited from time to time to double check the maturity levels.

7 Conclusion and Future Work

The discussion of how an overall score for a maturity level for security controls across different assets shows that the aggregation is an important tool needed to distinguish how the information security managers would optimize information security. In practice it makes a big difference which aggregation is used because it could lead to optimizing only the control maturity levels which are easily reachable. The defined priorization is necessary in order not to depend too much on the different kind of optimization strategies of the managers. This way, it can be steered more directly where the security should be enhanced and it probably also reflects better the current security level of companies. This approach is a helpful result for all companies aiming to regularly assess and improve their security as requested by the GDPR in order to ensure the confidentiality, integrity, availability and resilience of IT assets and evaluating the effectiveness of the technical and organizational measures for ensuring the security process.

As future work the outcome with other approaches could be compared to sen how the aggregation has changes the influence. Additionally, one might need to find other ways to prioritize the different controls, since in this case it was easy since it's one of the AHPs natural properties. Further investigations have to been carried out in order to clarify the validity of the control maturity levels because of the containing bias. Additional work could also be carried out to check validity of scope in order to measure any changes in the results after the metrics have been introduced.

References

1. Abbas Ahmed, R.K.: Security metrics and the risks: an overview. Int. J. Comput. Trends Technol. **41**(2), 106–112 (2016)

2. Abraham, S., Nair, S.: A predictive framework for cyber security analytics using attack graphs. Int. J. Comput. Netw. Commun. **7**(1), 1–17 (2015)
3. Ahmed, Y., Naqvi, S., Josephs, M.: Aggregation of security metrics for decision making: a reference architecture. In: ACM International Conference Proceeding Series (2018)
4. Anderson, R., et al.: Measuring the cost of cybercrime. In: Böhme, R. (ed.) The Economics of Information Security and Privacy, pp. 265–300. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-39498-0_12
5. Anderson, R., et al.: Measuring the changing cost of cybercrime our framework for analysing the costs of cybercrime. In: Workshop on the Economics of Information Security (WEIS), pp. 1–32 (2019)
6. Beck, A., Rass, S.: Using neural networks to aid CVSS risk aggregation - an empirically validated approach. J. Innov. Digit. Ecosyst. **3**(2), 148–154 (2016)
7. Bland, M.: Estimating mean and standard deviation from the sample size, three quartiles, minimum, and maximum. Int. J. Stat. Med. Res. **4**(1), 57–64 (2015)
8. Böhme, R.: Security metrics and security investment models. In: Echizen, I., Kunihiro, N., Sasaki, R. (eds.) IWSEC 2010. LNCS, vol. 6434, pp. 10–24. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-16825-3_2
9. Cheng, P., Wang, L., Jajodia, S., Singhal, A.: Aggregating CVSS base scores for semantics-rich network security metrics. In: Proceedings of the IEEE Symposium on Reliable Distributed Systems (2012)
10. Doane, D.P., Seward, L.E.: Applied Statistics in Business and Economics. McGraw-Hill Higher Education, New York (2016)
11. Gordon, L.A., Loeb, M.P.: The economics of information security investment. ACM Trans. Inf. Syst. Secur. **5**(4), 438–457 (2002)
12. Homer, J., et al.: Aggregating vulnerability metrics in enterprise networks using attack graphs. J. Comput. Secur. **21**(4), 561–597 (2013)
13. ISACA: COBIT 5: A business framework for governance and management of enterprise IT (2012)
14. ISO/IEC 27001: Information technology - security techniques - information security management systems - requirements. International Organization for Standardization (2013)
15. ISO/IEC 27701: Security techniques - extension to ISO/IEC 27001 and ISO/IEC 27002 for privacy information management - requirements and guidelines. International Organization for Standardization (2019)
16. Khajouei, H., Kazemi, M., Moosavirad, S.H.: Ranking information security controls by using fuzzy analytic hierarchy process. Inf. Syst. e-Bus. Manag. **15**(1), 1–19 (2017)
17. Lee, M.C.: Information security risk analysis methods and research trends: AHP and fuzzy comprehensive method. Int. J. Comput. Sci. Inf. Technol. (IJCSIT) **6**, 29–45 (2014)
18. Manikandan, S.: Measures of central tendency: median and mode. J. Pharmacol. Pharmacother. **2**(3), 214–215 (2011)
19. Nasser, A.A.: Measuring the information security maturity of enterprises under uncertainty using fuzzy AHP. I.J. Inf. Technol. Comput. Sci. **4**, 10–25 (2018)
20. Ramos, A., Lazar, M., Filho, R.H., Rodrigues, J.J.: Model-based quantitative network security metrics: a survey. IEEE Commun. Surv. Tutor. **19**(4), 2704–2734 (2017)
21. Rudolph, M., Schwarz, R.: Security indicators - a state of the art survey public report. FhG IESE VII(043) (2012)

22. Saleh, M.: Information security maturity model. Int. J. Comput. Sci. Secur. (IJCSS) **5**, 21 (2011)
23. Savola, R.M.: Towards a taxonomy for information security metrics. In: Proceedings of the ACM Conference on Computer and Communications Security, pp. 28–30 (2007)
24. Schmid, M., Pape, S.: A structured comparison of the corporate information security maturity level. In: Dhillon, G., Karlsson, F., Hedström, K., Zúquete, A. (eds.) SEC 2019. IAICT, vol. 562, pp. 223–237. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-22312-0_16
25. Schmitz, C., Pape, S.: LiSRA: lightweight security risk assessment for decision support in information security. Comput. Secur. **90** (2020)
26. Syamsuddin, I., Hwang, J.: The application of AHP to evaluate information security policy decision making. Int. J. Simul. Syst. Sci. Technol. **10**(4), 46–50 (2009)
27. Vinet, L., Zhedanov, A.: A 'missing' family of classical orthogonal polynomials. J. Phys. A Math. Theor. **44**(8), 16 (2011)

B.6 ESARA: A Framework for Enterprise Smartphone Apps Risk Assessment

Majid Hatamian, Sebastian Pape, and Kai Rannenberg. ESARA: A framework for enterprise smartphone apps risk assessment. In *ICT Systems Security and Privacy Protection - 34th IFIP TC 11 International Conference, SEC 2019, Lisbon, Portugal, June 25-27, 2019, Proceedings*, pages 165–179, 06 2019. doi: 10.1007/978-3-030-22312-0_12. URL https://doi.org/10.1007/978-3-030-22312-0_12

ESARA: A Framework for Enterprise Smartphone Apps Risk Assessment

Majid Hatamian[1(✉)], Sebastian Pape[1,2], and Kai Rannenberg[1]

[1] Chair of Mobile Business & Multilateral Security, Goethe University Frankfurt, Frankfurt, Germany
{majid.hatamian,sebastian.pape,kai.rannenberg}@m-chair.de
[2] Chair of Information Systems, University of Regensburg, Regensburg, Germany

Abstract. Protecting enterprise's confidential data and infrastructure against adversaries and unauthorized accesses has been always challenging. This gets even more critical when it comes to smartphones due to their mobile nature which enables them to have access to a wide range of sensitive information that can be misused. The crucial questions here are: How the employees can make sure the smartphone apps that they use are trustworthy? How can the enterprises check and validate the trustworthiness of apps being used within the enterprise network? What about the security and privacy aspects? Are the confidential information such as passwords, important documents, etc. are treated safely? Are the employees' installed apps monitoring/spying the enterprise environment? To answer these questions, we propose *Enterprise Smartphone Apps Risk Assessment (ESARA)* as a novel framework to support and enable enterprises to analyze and quantify the potential privacy and security risks associated with their employees' installed apps. Given an app, *ESARA* first conducts various analyses to characterize its vulnerabilities. Afterwards, it examines the app's behavior and overall privacy and security perceptions associated with it by applying natural language processing and machine learning techniques. The experimental results using app behavior and perception analyses indicate that: (1) *ESARA* is able to examine apps' behavior for potential invasive activities; and (2) the analyzed privacy and security perceptions by *ESARA* usually reveal interesting information corresponding to apps' behavior achieved with high accuracy.

Keywords: Smartphone · App · Security · Privacy · Risk · Enterprise

1 Introduction

The amount of available apps for smartphones seems to be almost endless. The developers range from spare time developers to large companies. However, none

The original version of this chapter was revised: An acknowledgement was added. The correction to this chapter is available at https://doi.org/10.1007/978-3-030-22312-0_27

Published by Springer Nature Switzerland AG 2019
G. Dhillon et al. (Eds.): SEC 2019, IFIP AICT 562, pp. 165–179, 2019.
https://doi.org/10.1007/978-3-030-22312-0_12

of the app stores offers a dedicated security or privacy score for those apps [19]. This does not only challenge individuals but also companies. However, smartphones are often used for personal matters and official business. *Bring your own device (BYOD)* is an attractive employee IT ownership model that enables employees to bring and use their personal devices in enterprises. Such a model provides more flexibility and productivity for the employees, but may impose some serious privacy and security risks. This way not an administrator decides about the installation of apps but the user. Similar problems arise if users are allowed to install apps on the devices provided by the company. The problem raises when enterprise's confidential data is endangered as smartphones now being used to access enterprise email, calendars, apps and data. As a result, enterprises are facing the tricky task of protecting valuable data from threats such as data leakage and malware. As a consequence, it is quite challenging for enterprises to balance both their employees' needs and their security concerns. But even if the employees are not allowed to decide by themselves, then the decision would have to be made by the IT department. As a consequence, enterprises would have to provide black lists that contain apps that are not allowed to be used, or white lists that contain apps allowed for use. Grey lists may be established to list apps, where no decision was made. In any case either the IT department needs to make decisions which app belongs to which list or the employees need to make their own decisions, whether a specific app is to be used. Decisions will be made as a trade off between the necessity of the app for business purposes and the risk with regard to enterprise assets.

Our Work: In this paper, we propose *Enterprise Smartphone Apps Risk Assessment (ESARA)* as a novel framework aimed at supporting enterprises to protect their data against adversaries and unauthorized accesses. Our framework eases the process of privacy and security risk assessment for the use of smartphone apps. To achieve this goal, we propose two concepts regarding the privacy and security assessment of smartphone apps namely app *Behavior Analyzer (BA)* and app *Perception Analyzer (PA)*. We develop these two concepts along with two essential requirements namely *vulnerability checker* and *malware checker* that are cooperated with each other aiming at supporting enterprises to discriminate privacy and security misbehaviors. To the best of our knowledge, we are the first proposing the combination of these concepts and requirements that are jointly working with each other. Through experiments and implementations, we investigate how efficient and reliable the newly proposed concepts are.

Outline: The rest of this paper is organized as follows. Section 2 reviews the existing works in the area of smartphone app privacy and security preservation for general and enterprise use cases. In Sect. 3 the respective components and architecture of *ESARA* framework are presented. Section 4 elaborates on the main results obtained from the evaluation of different components of *ESARA* and highlights the key insights. Finally, we present the main conclusions in Sect. 5.

2 Related Work

In this section, we provide an overview of the relevant related work in the area of privacy and security enhancement in smartphone ecosystems and enterprise environments.

Agarwall and Hall [14] propose an approach called *ProtectMyPrivacy (PMP)* for iOS devices to detect access to private information at run-time and protect users by substituting anonymized data to be sent instead of sensitive information. Enck et al. [17] proposed *TaintDroid* for real-time tracking of information flows of smartphone apps. By focusing on personal resources, the system can reveal the manipulation or transfer of sensitive data and thus analyze the app's behavior. The monitoring procedure is based on identifying privacy-related information sources and labeling associated data. Moreover, other impacted data are tracked and identified before being transferred outside the system. The evaluation on 20 popular apps showed data leakage, e.g. phone identifier, location information and phone number being transferred to remote advertising servers. The *Apex* mechanism [24] is a name for an additional component for Android which enables users to selectively allow, deny or limit access to specific permissions requested by apps. Beresford et al. [15] propose an approach called *Mockdroid* to substitute private data with mock data when they are asked to be accessed by installed apps. *TISSA* [27] is another component for Android that enables user to choose a list of untrusted apps, and based on this list it provides mock data in place of private data at run-time. Appicaptor [12] is a framework that helps enterprises for app risk management. The goal of Appicaptor is to detect the potential privacy and security risks associated with mobile app by benefiting from static analysis of app binaries. Based on app's behavior, a ranking list is provided to classify apps into white and black lists. BizzTrust [9] is another framework that suggested the use of restricted and open areas on the employee's smartphone. Both approaches are mainly focused on security risks resulted from malicious apps.

We believe that one efficient solution should not only be focused on security behavior of mobile apps, but also privacy behavior. Importantly, consideration of users' perception about the behavior of apps plays an important role to have a more comprehensive solution. These are interesting works, but neither of them focuses on a comprehensive solution that fulfills the essential requirements of enterprise environment. In our work, we propose a solution that enhances the existing works and revamps the current enterprise app risk assessment models.

3 ESARA Framework

3.1 Goal and Requirements

Our framework makes use of different approaches from literature and combines them with our app behavior analyzer and our app perception analyzer to get a more realistic and holistic picture of installed apps. E.g., a malware checker does not detect data leakages or vulnerabilities in an app and a vulnerability

scanner does not detect malicious behavior. Requirements for the development of *ESARA* were: (1) Reusing existing approaches; (2) Limiting the effort needed (since there is a large number of apps); (3) Scalability in the way that it should be easy to rely on external services and allowing several companies to share a same infrastructure; (4) Independence from app markets since even after several years none of them offers a decent security or privacy score; (5) Involving employees for feedback when using an app; (6) Involving employees for decisions.

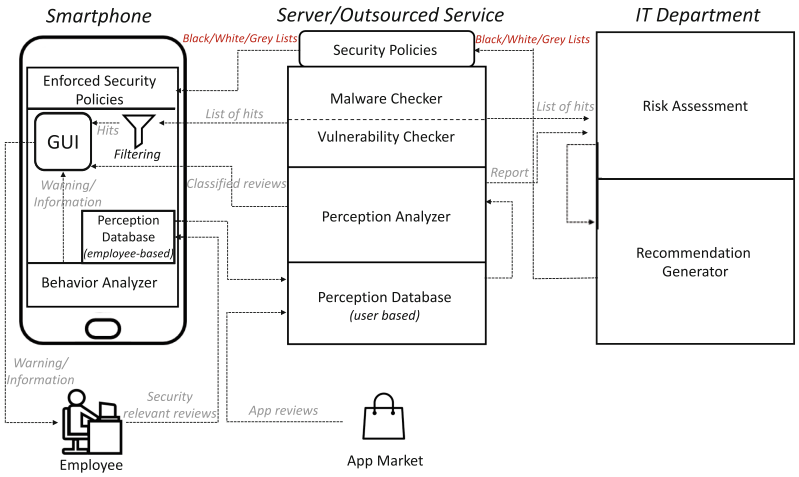
3.2 Architecture Design

Figure 1a shows an overview of the proposed architecture for *ESARA*. As can be seen, *ESARA* consists of three main modules: employee's smartphone, server and enterprise IT department. On the employee's device an app is running which analyzes the behavior of a certain installed app and ultimately communicates the results to the employee. It also stores the employee's security and privacy perception and receives results regarding the perception analysis and risk assessment from the other two modules. The server or an outsourced service is supposed to check apps for vulnerabilities and malicious activities by running a malware and vulnerability scanner, therefore, it does not collect any data from employees. This server/service is also responsible to analyze employees' and other users' perception about security and privacy behavior of apps. If security policies are put in place, black, white and gray lists can also be stored here. The enterprise IT department takes the final decisions about which app is to place on which list – either manually or automatically by defining certain rule sets.

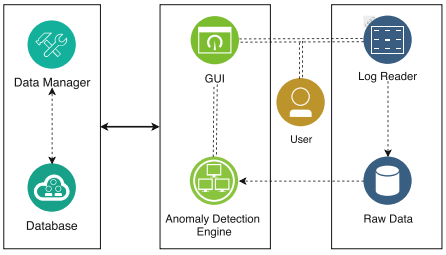
3.3 Components

Malware Checker. The impact of infection by a malware can be huge ranging from enterprise's infrastructure to the entire network. This component ensures the protection of enterprise's confidential data against malware. Therefore, we should not neglect the importance of this aspect of mobile apps while designing the *ESARA*'s architecture. As deploying a malware checker on resource constrained smartphones can be challenging [16], we propose the use of malware checker within the cloud as it has more computational resources. Therefore, this component is running on the server side. Checks need to be repeated with each update of the app or update of the malware checker's signature file.

Vulnerability Checker. Vulnerabilities are exploited by hackers to gain access to the device's or enterprise's resources. Statistics and observations showed that mobile platforms are among the most vulnerable operating systems in 2017 [13]. An observation by NowSecure [10] demonstrated that 25% of mobile apps have at least one high risk security vulnerability. Also, the latest security report published by Arxan [11] showed that 59% of the analyzed Android finance apps contained three OWASP mobile top 10 risks [3]. Surprisingly, all the analyzed



(a) ESARA



(b) Behavior analyzer

Fig. 1. High level overviews

iOS apps had at least 3 top risks. Due to such shocking statistics, an in-depth vulnerability analysis is required to investigate the potential vulnerabilities imposed by the employees' installed apps. Therefore, we also consider the importance of vulnerability analysis in *ESARA*'s architecture. Similar to the malware checker, this component is also running on the server side. There is an availability of a diverse number of vulnerability checkers both for Android and iOS that can be exploited based on the requirements [2,4–7,22].

Behavior Analyzer. *Behavior Analyzer (BA)* is an extension of our previous work [18,20] and a monitoring tool that analyzes the behavior of employee's installed apps. In contrast to run-time monitoring, where one could conclude what an employee was doing, we analyze the apps' behavior only by looking at the apps' permission requests. This way, the employees' privacy will be respected, while on the other hand security intrusive apps can be identified. Figure 1b shows

a high level architecture of the *BA* tool. In what follows, we elaborate on the core parts of *BA* and their respective role.

Log Reader. The log reader collects the logs from `AppOpsCommand` and it sends a timer to the `PermissionUsageLogger` service periodically. When it is received, the logger queries the AppOps service that is already running on the phone for a list of apps that have used any of the operations we are interested in tracking. We then check through that list and for any app that has used an operation more recently than we have checked, we store the time at which that operation was used. These timestamps are then counted to get a usage count.

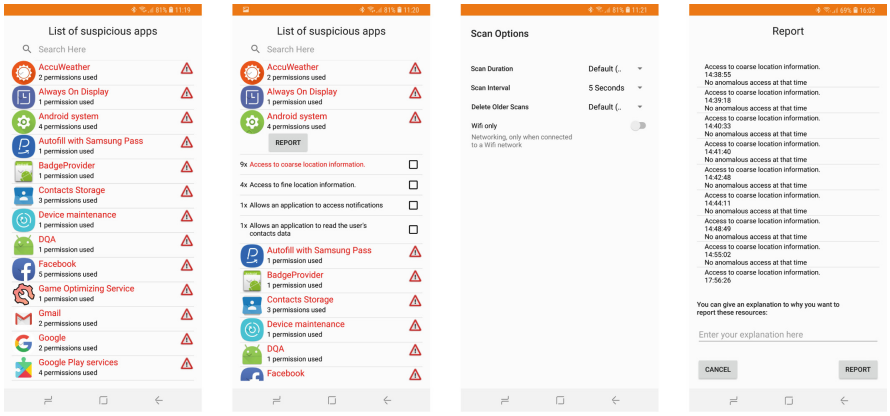
Anomaly Detection Engine. This component is supposed to behaviorally analyze the installed apps by getting help from the results obtained from the log reader component. This is done according to a rule-based mechanism which is supposed to increase the functionality and flexibility of our approach. Consequently, we have defined a set of invasive behavior detection rules that are aimed to analyze the behavior of employees' installed apps. We initially defined a set of sensitive permissions (introduced by Android[1]) and we mainly analyze the accesses to these resources. While implementing the *BA* tool, we paid special attention to the following elements to discern which resource access might be legitimate (needed by a certain app):

- Device's Orientation: This gives us information about the orientation of the device, e.g., if the screen is down or up;
- Screen State: It describes whether the device's screen is on or off at a certain time. As long as a scan is running, we register a `Receiver` for the events `ACTION_SCREEN_ON` and `ACTION_SCREEN_OFF`;
- Proximity Sensor: The screen state alone, however, is not meaningful enough, as it may happen that the screen is indeed off but certain personal resources may still be accessed (e.g., when talking on the phone, the screen turns off when the phone is approaching the ear, but access to `RECORD_AUDIO` is justified at this time). Therefore, we read the proximity sensor to indicate whether an object is within a defined range of the mobile phone;
- App State: We also consider the app state (at the time of access to a certain resource) as an important element while monitoring the apps' behavior. We distinguish the following app states: `SYSTEM_APP`, `PRE_INSTALLED_APP`, `INACTIVE`, `BACKGROUND` and `FOREGROUND`.

The *BA* tool operates like a watchdog and it only checks whether sensitive device resources are accessed. To protect the employees' privacy, the *BA* does not have the right/capability to access the sensitive data itself or track/monitor employee's activities. Furthermore, privacy controls are given to the employees to selectively choose the information that they want to share with the IT department. In particular, the IT department does not learn about all apps on the employees' device but only about those where the employees submit a report.

[1] https://developer.android.com/guide/topics/permissions/requesting.html.

For the sake of user interface design and risk indicator communication to the employees, we designed user interfaces for *BA* as shown in Fig. 2.

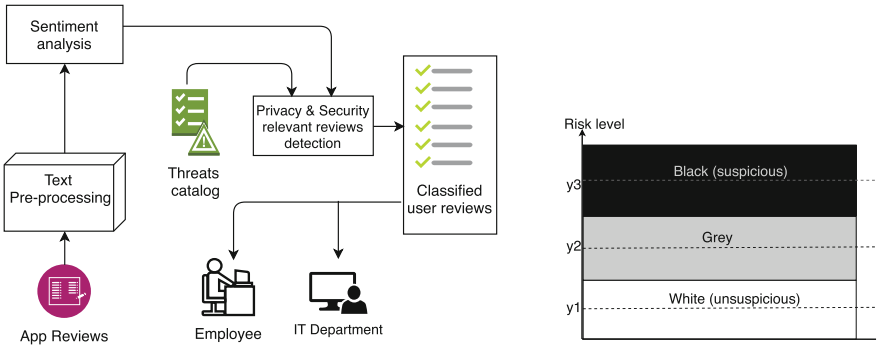


(a) List of Suspicious Apps (b) Details of Analysis (c) Analysis Settings (d) Report Option

Fig. 2. The proposed GUI for the BA tool

Perception Analyzer. The *BA* tool enables employees to write (optional) reviews regarding each privacy and security invasive activity that they observe. The main goal of *Perception Analyzer (PA)* as an extension of our previous work [21] is to mine these bunch of reviews to investigate how much privacy and security relevant claims/statements can be extracted that can be ultimately used for the risk assessment component. These self-written reports are sent to the IT security department of the enterprise as well. The main idea is to not only rely on individual's report, but also to consider a high level overview of apps' real behavior. This would enable the enterprise to improve the fairness of their decisions. Additionally, we enriched the reviews of the employees with reviews from app markets (e.g. Google Play). The reviews in app markets are in general more concerned about features and performance of the apps and only little of them contain comments about security or privacy. However, since for some of the apps there are tons of reviews, even a low percentage of reviews dealing with security and privacy can be helpful. Therefore, we used a machine learning approach to find the relevant reviews. Figure 3a shows the proposed architecture for *PA*.

The *app reviews* are first pre-processed in *text pre-processing* component using typical natural language processing (NLP) techniques (e.g. tokenization, stemming and removing stop words). Further, we propose the use of *sentiment analysis* techniques to find both positive and negative reviews that talk about privacy and security aspects of apps. Afterwards, the machine learning model



(a) Privacy and Security Reviews Analyzer (b) Risk Assessment in *ESARA*

Fig. 3. Review analyzer and risk assessment

comes in and *threat catalog* helps to identify the associated threats with each user review by getting help from *privacy and security relevant reviews detection*. Finally the *classified reviews* are communicated to the *IT department* for risk assessment procedure. As it is obvious, *BA* is supposed to tell the IT security department how good/bad is a certain app in terms of privacy and security aspects based on its behavior in reality, and *PA* is aimed at providing a fair comparison by considering a consensus from employees and crowdsources. We detect not only a privacy and security relevant user review, but also determine the threat hidden in it. To this end, we take the most relevant threats in the context of smartphone ecosystems introduced in [21] into account. These threats are used as the input for the supervised classification algorithm as described in Table 1.

Table 1. Identified threats

#	Threat	Description
T1	Tracking & Spyware	Allows an attacker to access or infer personal data to use it for marketing purposes, such as profiling or targeted ads
T2	Phishing	An attacker collects user credentials (e.g. passwords and credit card numbers) by means of fake apps or messages that seem genuine
T3	Unauthorized charges	The hidden and unauthorized charges through registration to a premium service AND/OR installation a certain app
T4	Unintended data disclosure	Users are not always aware of all the functionality of smartphone apps. Even if they have given explicit consent, users may be unaware that an app collects and publishes personal data
T5	Targeted ads	Refers to unwanted ads and push notifications
T6	Spam	Threat of receiving unsolicited, undesired or illegal messages. Spam is considered an invasion of privacy. The receipt of spam can also be considered a violation of our right to determine for ourselves when, how, and to what extent information about us is used
T7	General	Comprises all the threats that are not categorized into other categories, e.g. permission hungry apps, general security concerns, etc.

Risk Assessment. Risk assessment examines the potential privacy and security risks associated with each employee's installed app. Therefore, it is highly dependent on the results obtained from *BA*, *PA*, *malware checker* and *vulnerability checker*. In this paper, we assume three different risk levels, including black (seems suspicious), grey (requires more investigation) and white (seems unsuspicious) as shown by Fig. 3. If a certain app does not successfully pass the investigations done by malware and vulnerability checkers, then it is automatically ranked as black and the outcome will be communicated to the employee. Otherwise, the risk assessment considers the real behavior and overall perception results in order to provide the recommendation generator with sufficient decision making information. It is worth mentioning that our main focus is on the grey risk level.

Recommendation Generator. Recommendation generator gets the input from the risk assessment component. It helps the IT security departments to better classify apps as allowed or not allowed (e.g. blacklists and whitelists). Thus, it ranks similar functionality apps, i.e. those apps that have similar functionality (e.g. weather forecasting apps, navigation apps, etc.) are assigned ranks based on the analysis done by risk assessment. Moreover, it maintains a history of privacy and security behavior records (analyses) based on apps' versions, meaning that once a certain installed app is updated, a trend containing the behavior measurements related to the current and older versions will be issued. Therefore, the IT security department can follow and analyze the trend analyses done by *ESARA* from version to version. This would enable them to analyze the behavior of current versions and compare it with older versions. This is a substantial impact of *ESARA* that is based on the fact that there is no guarantee for privacy and security friendly apps to behave nicely in the future.

4 Evaluation

Since the efficiency of malware and vulnerability scanner is a topic of its own, we do not discuss it here. However, we discuss the results of the app behavior analyzer and the app perception analyzer. For the overall evaluation, we will discuss which component covers which kind of risk.

4.1 App Behavior Analysis Results

To evaluate the applicability of *BA* and its importance in the overall performance of *ESARA*, we demonstrate some initial results regarding the behavior analysis done by *BA*. To make our scope as narrow as possible and to have a fair analysis, we mainly focused on Android apps and chose one general purpose app category. To this end, we found *Health & Fitness* as the most interesting option that is widely used by people and has raised serious privacy and security concerns [23,26]. Therefore, we selected the top 20 apps in the *Health & Fitness* category and started the case study. We purchased six Android smartphones and installed

all the 20 apps on each of them. We then used *BA* to analyze the behavior of the aforementioned apps. While we were implementing the case study, the *BA* tool was running in the background the whole time (i.e. it was monitoring apps' behavior). We ran the apps once and let them to be executed in the background. Thus, we never interacted with the mobile devices during the experiment period. This is mainly because mobile apps are task-specific and expected to only access resources when needed for their functionality. When the employee is using an app, it is harder to infer whether the app needs to have access to a certain sensitive resource as this requires to know what exactly the employee is doing which may violate his/her privacy. But when the app is not used, it is easy to detect non-security friendly sensitive resource accesses, e.g. access to enterprise's confidential data (e.g. employees' calendar, contacts, ...), since most of them will be unsolicited. Afterwards, we collected and analyzed the data generated by *BA*. In total, nine sensitive resources were accessed by the apps. The results of the analysis for each app and resources are shown in Table 2. The numbers in each cell show the number of times that each app accessed a certain resource.

Table 2. Resource access behavior pattern extracted by *BA*.

Resources	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
STORAGE	561	106	143	106	175	103	124	196	702	186	53	394	137	87	156	747	95	331	184	1376
CAMERA	0	0	0	0	0	0	0	25	53	86	0	0	0	0	0	0	0	15	0	14
READ_SMS	0	0	0	0	0	0	0	0	17	0	0	0	0	0	0	0	0	0	0	0
READ_CONTACTS	0	62	0	0	0	0	0	53	31	653	0	0	34	0	0	0	0	0	0	142
LOCATION	0	32	0	0	0	0	985	183	650	403	217	0	116	96	412	3780	0	670	566	1526
PHONE_STATE	0	0	0	35	0	0	284	0	534	87	0	0	0	0	0	0	0	0	0	0
MICROPHONE	0	0	0	0	0	0	0	0	0	0	0	0	0	552	0	0	34	0	0	0
GET_ACCOUNTS	0	126	0	0	0	0	93	407	279	363	0	0	0	0	0	0	0	0	101	455
BODY_SENSOR	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

The accesses to `READ_STORAGE` are not surprising because the smartphones were not completely turned off and all apps could read and write files placed on the external storage (e.g. cache files). However, five apps accessed `CAMERA` (apps 8, 9, 10, 18 and 20). These accesses are not privacy-friendly, since the user does not know that the app currently accesses the camera. Furthermore, `READ_CONTACTS` was accessed by six apps. In general, such accesses to the contacts should not be done by apps. In our case the apps are health-based, where it is not clear why they need access to the user's contacts. `PHONE_STATE` is an interesting data resource since the respective information is highly sensitive. This permission enables an invasive party to gain access to sensitive resources such as phone number, cellular network information, outgoing call information, etc. The only relevant reason to access this permission is to stop the app when there is an ongoing call, however, we did not use SIM card on the devices, therefore, there is no reason of such resource access. This also happened to other sensitive resources such as `MICROPHONE`, `READ_SMS`, `LOCATION`, etc. We also observed that many of these resource accesses happened when: (1) the devices were in horizontal

orientation, (2) the devices' screen was off and (3) the proximity sensor indicated that there is no nearby object.

4.2 App Perception Analysis Results

To validate the capability of our novel perception analyzer, we collected a dataset consisting of 75,601 user reviews corresponding to these 20 health-based apps using the scraper in [1]. Three experts went manually through the data and labeled them. We then used $CountVectorizer$ and $TfidfTransformer$ packages in scikit-learn [25] for the feature extraction phase. We then split the data set into training and testing data (70% for training and 30% for testing). Using scikit-learn we exploited several classification algorithms such as *Support Vector Machines (SVMs), Random Forest, Logistic Regression (LR),* etc. We observed *LR* outperforms others, therefore, we only show the results for *LR*. We used recall, precision and F-score metrics to evaluate the performance of the classifier. The values of these metrics show how well the classifier's results correspond to the annotated results. Table 3 shows the values for the aforementioned metrics corresponding to each identified threat. The observation is that the overall recall and precision values are of 88.95% and of 91.16%, respectively. Moreover, the values obtained for F-score show the good performance of our approach.

Table 3. Performance measures of the classification algorithm

Classes	Recall	Precision	F-score
Tracking & Spyware	0.7549	0.8311	0.8214
Phishing	0.8588	0.8653	0.8601
Unauthorized charges	0.7912	0.9583	0.8296
Unintended data disclosure	0.9010	0.9765	0.9218
Targeted ads	0.9374	0.9971	0.9663
Spam	0.9374	0.9514	0.9388
General	0.7576	0.8639	0.8492
Overall	0.8895	0.9116	0.9059

Table 4 shows some examples regarding the strength of perception analyzer in distinguishing different types of user reviews with different sentiments and relevant threat (shown by T). The obtained results clearly confirm the applicability and the positive influence of perception analyzer in the overall risk assessment done by *ESARA*.

4.3 Risk Coverage

As *ESARA* is a privacy and security risk assessment tool for mobile apps in enterprises, it is of particular importance to check its coverage of the most prevalent

Table 4. An example of classified user reviews

#	Sample user review	T
1	*You don't need to spy on my activities outside of this app. they don't care about their customers, they want to ruin the device with horrible bloatware spyware*	T1
2	*Im still getting warnings that my phone is infected with virus after i update and scan again. If its not going to work why download it. I have very limited memory to use. No need to download stupid apps that dont work*	T2
3	*Cheating Y the hell.. u cut my 50 rupees for nothing.. i just enter my card details and u cut my money without asking me.. i want it back*	T3
4	*SHit!Takes control of device.. why my photo is there??!!*	T4
5	*Ads are terrible Sorry but the ads are comparing to the website really irritating.*	T5
6	*Had this problem about these Annoying full screen PoP-ups!*	T6
7	*Dangerous! requires unnecessary access to sensitive permissions! Uninstalled*	T7

mobile app risks. We took Veracode [8] as one of the well-established references that categorizes the top 10 mobile app risks (considering the top 10 risks introduced by OWASP [3]) and we investigate the robustness of *ESARA* in assessment and detection of each individual risk. In Table 5 we clarify which component of *ESARA* may detect which identified risk. Thanks to the novel combination of *BA*, *PA*, malware and vulnerability checkers, the *ESARA*'s components totally (shown by ✓) or partially (shown by (✓)) cover all the risks. We observed that each risk is at least covered by two components (except UI impersonation which is one the most complex risk scenarios in terms of identification and mitigation).

Table 5. Coverage of Veracode top 10 mobile app [8] risks by *ESARA*.

No.	Risk	Malware checker	Vuln. checker	Behavior analyzer	Perception analyzer
1	Activity monitoring and data retrieval	✓	–	(✓)	(✓)
2	Unauthorized dialing, SMS, and payments	✓	–	✓	(✓)
3	Unauthorized network connectivity	(✓)	–	–	(✓)
4	UI Impersonation	–	–	–	(✓)
5	System modification	✓	–	–	✓
6	Logic or Time bomb	✓	(✓)	–	–
7	Sensitive data leakage	(✓)	✓	✓	✓
8	Unsafe sensitive data storage	–	✓	–	(✓)
9	Unsafe sensitive data transmission	–	✓	–	✓
10	Hardcoded password/keys	✓	✓	–	–

4.4 Discussion and Limitations

We could address all the requirements we defined in Sect. 3.1 and cover the top 10 mobile app risks with at least two components (except for UI impersonation). The results from the evaluation of the app behavior analyzer and the app perception analyzer are very promising. However, there is a limitation of our work. We have not tested our framework in a real company environment, yet. We only did user studies in a laboratory environment. Therefore, it remains to respectively show that employees would like the idea of getting support for the decisions about which apps they want to install and therefore actively make use of the potentials provided by *ESARA*.

5 Conclusion and Future Work

Smartphones have become ubiquitous within enterprise environments. At the same time, with the increased interest and not only the adoption of *BYOD*, employees heavily rely on apps, sometimes also used for personal purposes, that have access to enterprise confidential data as well. As a result, security and privacy have become a big challenge in enterprises. In this paper, we proposed *ESARA* as a novel framework to analyze and quantify the potential privacy and security risks associated with employees' smartphone apps within an enterprise environment. After an in-depth analysis of the most relevant works in the literature, we proposed an approach that leverages a four-pillar mechanism, including malware checker, vulnerability checker, behavior analyzer and perception analyzer. The combination of these mechanisms that are jointly working together supports and enables enterprises to profoundly examine the privacy and security aspects of their employees' installed apps. Since malware and vulnerability checkers are well researched, we only evaluated the performance of the two newer components, the behavior analyzer (*BA*) and the perception analyzer (*PA*). We practically showed the applicability of using behavior and perception analyses to have a more fine-grained app risk assessment and our results confirmed that these two factors play a critical role in the overall quantification of app security and privacy risks. *ESARA* opens opportunities for further innovative solutions for risk assessment of mobile apps within enterprise environments, including easing the quantification of apps trustworthiness degree.

In our future work, we will further enhance the performance of the perception analysis component by providing more training and testing data. Additionally, user studies are planned to determine the employees' and IT departments' acceptance of our approach. Also, a comprehensive analysis in an enterprise environment to validate the whole framework in a real world scenario is planned in the future.

Acknowledgment. This research was supported by the European Union's Horizon 2020 Research and Innovation program under the Marie Skłodowska-Curie "Privacy&Us" project (GA No. 675730).

References

1. Google play scraper. https://github.com/facundoolano/google-play-scraper/
2. Mobile application security scanner. https://www.ostorlab.co/
3. Mobile top 10 2016-top 10. https://www.owasp.org/index.php/mobile_top_10_2016-top_10/
4. Nviso. apkscan. https://apkscan.nviso.be/
5. Quick android review kit. https://github.com/linkedin/qark
6. Quixxi integrated app management system. https://quixxisecurity.com/
7. Sanddroid - an automatic android application analysis system. http://sanddroid.xjtu.edu.cn
8. Veracode mobile app top 10. http://www.veracode.com/directory/mobileapp-top-10/
9. Protection of sensitive data and services (2012). https://www.sit.fraunhofer.de/en/bizztrust/
10. NowSecure mobile security report (2016). https://www.nowsecure.com/blog/2016/02/11/2016-nowsecure-mobile-security-report-now-available/
11. Arxan's 5th annual state of application security report (2016). https://www.arxan.com/press-releases/arxans-5th-annual-state-of-application-security-report-reveals-disparity-between-mobile-app-security-perception-and-reality
12. Framework for app security tests (2016). https://www.sit.fraunhofer.de/en/appicaptor/
13. Most vulnerable os of the year 2017 (2017). https://www.cybrnow.com/10-most-vulnerable-os-of-2017/
14. Agarwal, Y., Hall, M.: Protectmyprivacy: detecting and mitigating privacy leaks on IOs devices using crowdsourcing. In: Proceedings of MobiSys, pp. 97–110 (2013)
15. Beresford, A., Rice, A., Sohan, N.: Mockdroid: trading privacy for application functionality on smartphones. In: The Proceedings of the 12th Workshop on Mobile Computing Systems and Applications, Phoenix, Arizona, USA, pp. 49–54 (2011)
16. Chandramohan, M., Tan, H.B.K.: Detection of mobile malware in the wild. Computer **45**(9), 65–71 (2012). https://doi.org/10.1109/MC.2012.36
17. Enck, W., et al.: Taintdroid: an information-flow tracking system for realtime privacy monitoring on smartphones. In: The Proceedings of the the 9th ACM USENIX Conference on Operating Systems Design and Implementation, Vancouver, BC, Canada, pp. 393–407 (2010)
18. Hatamian, M., Serna, J., Rannenberg, K., Igler, B.: Fair: fuzzy alarming index rule for privacy analysis in smartphone apps. In: The Proceedings of the 14th International Conference on Trust and Privacy in Digital Business (TrustBus), Lyon, France, pp. 3–18 (2017)
19. Hatamian, M., Serna-Olvera, J.: Beacon alarming: Informed decision-making supporter and privacy risk analyser in smartphone applications. In: Proceedings of the 35th IEEE International Conference on Consumer Electronics (ICCE), USA (2017)
20. Hatamian, M., Kitkowska, A., Korunovska, J., Kirrane, S.: "It's Shocking!": analysing the impact and reactions to the A3: Android Apps behaviour analyser. In: Kerschbaum, F., Paraboschi, S. (eds.) DBSec 2018. LNCS, vol. 10980, pp. 198–215. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-95729-6_13
21. Hatamian, M., Serna, J., Rannenberg, K.: Revealing the unrevealed: mining smartphone users privacy perception on app markets. Comput. Secur. (2019). https://doi.org/10.1016/j.cose.2019.02.010. http://www.sciencedirect.com/science/article/pii/S0167404818313051

22. Maggi, F., Valdi, A., Zanero, S.: Andrototal: a flexible, scalable toolbox and service for testing mobile malware detectors. In: Proceedings of the 3rd ACM Workshop on Security and Privacy in Smartphones and Mobile Devices, pp. 49–54 (2013)
23. Martínez-Pérez, B., De La Torre-Díez, I., López-Coronado, M.: Privacy and security in mobile health apps: a review and recommendations. J. Med. Syst. **39**(1), 1–8 (2015)
24. Nauman, M., Khan, S., Zhang, X.: Apex: extending android permission model and enforcement with user-defined runtime constraints. In: Proceedings of the 5th ACM Symposium on Information, Computer and Communications Security, pp. 328–332 (2010)
25. Pedregosa, F., et al.: Scikit-learn: machine learning in python. J. Mach. Learn. Res. **12**, 2825–2830 (2011)
26. Plachkinova, M., Andres, S., Chatterjee, S.: A taxonomy of mhealth apps - security and privacy concerns. In: 2015 48th HICSS, pp. 3187–3196, January 2015
27. Zhou, Y., Zhang, X., Jiang, X., Freech, V.W.: Taming information-stealing smartphone applications (on android). In: the Proceedings of the 4th International Conference on Trust and Trustworthy Computing, Pittsburgh, PA, USA, pp. 39–107 (2011)

B.7 An Insight into Decisive Factors in Cloud Provider Selection with a Focus on Security

Sebastian Pape and Jelena Stankovic. An insight into decisive factors in cloud provider selection with a focus on security. In *Computer Security - ESORICS 2019 International Workshops, CyberICPS, SECPRE, SPOSE, ADIoT, Luxembourg City, Luxembourg, September 26-27, 2019, Revised Selected Papers*, volume 11980 of *Lecture Notes in Computer Science*, pages 287–306, Cham, 09 2019. Springer International Publishing. ISBN 978-3-030-42048-2. doi: 10.1007/978-3-030-42048-2_19. URL https://link.springer.com/chapter/10.1007/978-3-030-42048-2_19

An Insight into Decisive Factors in Cloud Provider Selection with a Focus on Security

Sebastian Pape(✉) and Jelena Stankovic

Chair of Mobile Business and Multilateral Security,
Goethe University Frankfurt, Frankfurt, Germany
sebastian.pape@m-chair.de

Abstract. In the last ten years cloud computing has developed from a buzz word to the new computing paradigm on a global scale. Computing power or storage capacity can be bought and consumed flexibly and on-demand, which opens up new opportunities for cost-saving and data processing. However, it also goes with security concerns as it represents a form of IT outsourcing. We investigate how these concerns manifest as a decisive factor in cloud provider selection by interviews with eight practitioners from German companies. As only a moderate interest is discovered, it is further examined why this is the case. Additionally, we compared the results from a systematic literature survey on cloud security assurance to cloud customers' verification of their providers' security measures. This paper provides a qualitative in-depth examination of companies' attitudes towards security in the cloud. The results of the analysed sample show that security is not necessarily decisive in cloud provider selection. Nevertheless, providers are required to guarantee security and comply. Traditional forms of assurance techniques play a role in assessing cloud providers and verifying their security measures. Moreover, compliance is identified as a strong driver to pursue security and assurance.

Keywords: Cloud provider selection · Security assurance · Interviews

1 Introduction

Cloud Computing has been emerging as the new computing paradigm in the last ten years, enabling consumers to purchase computing power and storage capacity on-demand, conveniently and cost efficiently from specialized providers. Recent studies claim that cloud computing has left the hype phase behind and can already be considered the norm for IT [10].

Besides the potential economic benefits of cloud adoption, it also goes with security concerns as it represents a form of IT outsourcing and exhibits technological peculiarities concerning size, structure and geographical dispersion [35]. With rising adoption rates of cloud services, security concerns remained

S. Katsikas et al. (Eds.): ESORICS 2019 Workshops, LNCS 11980, pp. 287–306, 2020.
https://doi.org/10.1007/978-3-030-42048-2_19

unchanged or even rose as well. On the other hand, many technical reports also reveal benefits to security in the cloud. It is argued that a cloud provider (CP) enjoys economies of scale in terms of security as well, being able to invest more and thereby achieve a higher security level on a much larger scale than most client companies would with an in-house data centre [24,29]. Thus, in either case, one would expect companies to incorporate security into their provider selection and cloud use.

We investigate organizations' practises when selecting a secure CP: *"What role does security play in CP selection?"*. Despite expected "inherent differences in such things as the intended purpose, assets held, legal obligations, exposure to the public, threats faced, and tolerance to risk" between different companies or organizations [29], we expected to verify the importance of security. Under that assumption there would be an incentive for providers to invest in security measures, as potential customers might make their choice based on this characteristic [24]. Moreover, in order to prevent a market for lemons in cloud computing [1], we expected cloud service providers and customers to come up with quality/security assurance methods. Thus, we intended the follow-up question: *How are the providers' security measures verified?* – if security is a selection criteria. Or respectively: *Why is security not considered in CP selection?*

In order to find answers for the underlying research questions a qualitative approach is taken. Practitioners from eight German companies who are associated with CP selection are interviewed and questioned about their companies' provider selection and ways to establish assurance.

2 Related Work

Our research questions can be related to contributions on provider selection, the role of security and security assurance. Security concerns, which are seen as the inhibiting factor of cloud adoption, can be easily related to well researched issues. A bunch of issues is related to technical properties of cloud computing, i.e. the complex architecture [29], multi-tenancy in connection with isolation failures [24,29], and network vulnerabilities The list of risks also includes the threat of a malicious insider on the CP's side [9], who may abuse his privileges. However, this is a general outsourcing issues due to a loss of governance which can bear dangers for the cloud customers [24]. Therefore, focus in this section is on measures for the CP to assure the security level of its service (corresponding to our extended research question). Assurance is also often necessary from a legal and compliance perspective since most companies underlie a variety of legal obligations, depending on the sector and the type of data they handle.

Since we follow the qualitative content analysis method which is considered hermeneutic and uses deductive examination (cf. Sect. 3.2), an inherent understanding of the topic was necessary in order to interpret the material. Therefore, we conducted a systematic literature survey on security assurance measures.

Table 1. Reviewed contributions

Assurance	Contribution	Model proposals
SLAs	Lee et al. [40], Luna et al. [44]	Casola et al. [11], Kaaniche et al. [31], Nugraha and Martin [53]
Monitoring	Ismail et al. [27]	Ba et al. [8], Deng et al. [17], Fernando et al. [21], Kanstrén et al. [32], Rios et al. [62], Zhang et al. [71,72]
Testing		Sotiriadis et al. [67], Stephanow and Khajehmoogahi [68], Tung et al. [70]
Auditing	Ryoo et al. [64]	Ghutugade and Patil [22], Jakhotia et al. [28], Jiang et al. [30], Lins et al. [42,43], Ma et al. [45], Majumdar et al. [47], Meera and Geethakumari [48], More and Chaudhari [50], Parasuraman et al. [55], Pasquier et al. [56], Rashmi and Sangve [59], Rewadkar and Ghatage [61], Thendral and Valliyammai [69]
Certification	Di Giulio et al. [18], Di Giulio et al. [19], Polash and Shiva [57], Schneider et al. [65]	Anisetti et al. [3–5], Anisetti et al. [6], Katopodis et al. [33], Krotsiani and Spanoudakis [34], Lins et al. [41], Munoz and Mafia [51]
Other		Henze et al. [25], Mohammed and Pathan [49], Ramokapane et al. [58], Rizvi et al. [63], Sen and Madria [66]

2.1 Security Assurance

We rely on a survey from Ardagna et al. [7] which covers contributions on security measures and assurance techniques until 2014 and followed their methods and definitions as close as possible to update it for our recent research. Due to space limitations, we can not show the results in detail, but only give a brief summary and list them in Table 1.

Almost all contributions reasoned with customers' security concerns as the main inhibiting factor of cloud adoption and that a contribution might provide the needed transparency to resolve that issue. A further justification for new contributions on security assurance were the "special properties" of the cloud which raise new requirements for that topic. Clearly each contribution presented the benefits of its solution, some also covered the challenges, but the drawbacks of certain assurance techniques could only be found in a few contributions from adjacent categories. Certification and security SLAs were presented as the more accessible and convenient measures. In these contributions the customer is clearly involved in the negotiation and provider choice. On the contrary, contributions on auditing, monitoring and testing are mostly technical models or frameworks.

It might be difficult to apply these technical models and is it not clear if they are practical in reality and who would implement them.

2.2 CP Selection

In this section qualitative research which determined relevant criteria for CP selection will be discussed. The presented contributions suggest a formal and systematic selection process of a CP and identify security as a relevant criterion. They pursue similar research questions and use a qualitative approach like we do. Nevertheless, their results are narrowed down into compact lists, where security is identified as a requirement but not further discussed. We aim to close this gap, by giving further insight into experts' answers and the role of security.

Repschläger et al. [60] develop a CP classification model with a focus on infrastructure as a service (IaaS). The relevant target dimensions are determined as a result of expert interviews and validated and expanded through a literature review. The authors conduct five interviews with experts providing different perspectives on common objectives in cloud computing.

Similarly, Hetzenecker et al. [26] derive a model of requirements to support the user in evaluating CPs. Their model consists of six categories with in total 41 requirements. "Information security" is derived as a category with 15 requirements, such as integrity, availability, data disposal, encryption or scalability. All requirements are only presented by a title but not further elaborated.

Lang et al. [39] conduct a Delphi study with 19 decision makers in order to determine relevant selection criteria with a high abstraction level. Security is only identified as a component of the highest rated criterion "functionality" which does not permit to make any statements about the importance of security at all. The authors call for further research to investigate their identified requirements on a lower abstraction level.

2.3 Security, Threat Models and Compliance

Following the CSA top threats to cloud computing [12–15] as shown in Table 2 one can see that most of the threats are related to security and that data breaches soon evolve as the top threat. In an extensive survey Kumar and Goyal [37] map the threats also to requirements, vulnerabilities and countermeasures. Alhenaki et al. [2] investigate some of the threats mentioned by the CSA, do also a mapping to countermeasures and additionally identify the relevant cloud service models (Saas, PaaS, IaaS) which are concerned by the threats. Mahesh et al. [46] elaborate aspects of cloud computing that need special attention, i.e. by audits. They also list most prominent frameworks and working groups that are widely accepted across industries and describe some approaches from industry practices.

3 Methodology

In this section we briefly describe how the interviews were conducted and how the data was analysed.

Table 2. Top threats to cloud computing identified by CSA [12–15]

#	2010	2013	2016	2019
1	Abuse and nefarious use of cloud computing	Data breaches	Data breaches	Data breaches
2	Insecure application programming interfaces	Data loss	Weak identity, credential and access management	Misconfiguration and inadequate change control
3	Malicious insiders	Account hijacking	Insecure APIs	Lack of cloud security architecture and strategy
4	Shared technology vulnerabilities	Insecure APIs	System and application vulnerabilities	Insufficient identity, credential, access and key management
5	Data loss/leakage	Denial of service	Account hijacking	Account hijacking
6	Account, service & traffic hijacking	Malicious insiders	Malicious insiders	Insider threat
7	Unknown risk profile	Abuse of cloud services	Advanced persistent threats (APTs)	Insecure interfaces and APIs
8	–	Insufficient due diligence	Data loss	Weak control plane
9	–	Shared technology issues	Insufficient due diligence	Metastructure and applistructure failures
10	–	–	Abuse and nefarious use of cloud services	Limited cloud usage visibility
11	–	–	Denial of service	Abuse and nefarious use of cloud services
12	–	–	Shared technology issues	–

3.1 Sample Selection and Conduction of Interviews

We conducted semi-structured interviews with practitioners engaged in the selection of a CP, e.g. with the role of network or cloud architect or a management position. With semi-structured interviews we were able to get answers to a set of predetermined questions but were still flexible enough to include spontaneous questions arising from the discussion with the practitioners.

Since we could not offer financial compensation, we tried to get in touch with relevant practitioners at the Cloud Expo Europe 2018 and completed the set of interviewees with contacts from our personal network. The process of the invitation and the interviews was as follows: When inviting the participants, we already included the information that we were looking for experts in the field of cloud computing to find out which criteria were considered when choosing a CP and which requirements were imposed on the provider. Ideally, the participants should either be involved in such a decision. In order to be able to verify security as a criterion without revealing it beforehand, the research focus on security was not given in the invitation.

We first conducted a pilot interview to test and validate the interview guidelines. Respondents Ra and Rb were from the financial sector and related secu-

Table 3. Respondents' profiles

Respondents	Relation to the cloud	Sector	Employees	Expert's position
Ra/Rb	User	Financial Services	>1000	Infrastructure Specialists
R1	Consultant	IT Consulting	>100000	Cloud Advisory Sen. Manager
R2	Provider	IT	<50	CEO
R3	User	Financial Services	>10000	Network Architect
R4	User	Energy Supply	>10000	Cloud Architect
R5	User	Automotive	>100000	Solution Architect
R6	User	Financial Services	>1000	IT Security Manager
R7	User	Metal Processing	>1000	Project Manager (IT Infrastr.)
R8	User	Fintech	<50	CTO

rity closely to compliance, i.e. regulations imposed by the national supervisory authority BaFin. Therefore, the remaining interviews were further enriched by the question whether there was the intrinsic motivation or personal responsibility to select a secure provider. Afterwards, from October to December 2018, we interviewed eight respondents (cf. Table 3) face to face and in German. In order to maintain continuity all interviews were conducted by the same interviewer . Interviews had an average duration of around 37 min.

Due to space limitations, we describe the interview guideline only briefly. After the warm-up, the second block of questions addressed the provider selection. According to the research questions if respondents claimed to consider security when selecting a CP they were asked about possible assurance techniques their company used. In case security was not mentioned, the respondents were asked about the importance of security. Although security was not among the first criteria mentioned, it was present in most discussions. Eventually this lead to covering both sides of the decision tree in most of the interviews. Finally, the transparency on the cloud market was addressed to generate additional ideas for possible improvements to a non-transparent market.

3.2 Data Analysis

The interviews were transcribed word by word and analyzed with MAXQDA following the qualitative content analysis method from Kuckartz [36], since it suited the data collected in the semi-structured interviews and allowed to analyze the data with regard to the research questions. To get well acquainted with

Table 4. Coding frame for assurance techniques

Assurance techniques	Respondents talk about how they establish security assurance
Certification	Respondents talk about certification. The topic is either which ones they consider important or the advantages and drawbacks of certificates
Audits	Respondents audit their providers or talk about auditing. Statements are also included if they are about financial auditing
Contractual agreements	User and provider agree contractually on certain requirements the provider has to fulfill or on the right of the user to audit
Data center visits	Respondents place a value on being allowed to visit the provider's data center
Documentation	Respondents place a value on checking the providers' documentation on processes or technical measures
Penetration tests	The respondents run penetration tests as a mean of assurance
Cloud risk process	Companies' own process for risk assessment
Questionnaire on security measures	A company uses a questionnaire (comparable to CSA's CAIQ) in order to obtain information from a provider
Skepticism	Respondents express skepticism towards some assurance techniques, or the sense of assurance in general

the material, in the first phase of analysis each interview was summarized and the peculiarities of the given answers were noted. Next, master-codes were developed and tested on the first three interviews before coding the whole material. These codes were generated mostly deductively out of the interview questions. For instance, the codes "Provider Selection" and "Assurance Techniques" were rather straight forward, as these where the main research questions. The result of this phase was a list of master-codes. After coding the whole material with the master-codes, all passages coded with the same master-code were grouped and reread. At this point the aim was to differentiate the master-codes by inductively deriving sub-codes for each master-code. While proceeding from one interview to the next, the generated sub-codes were revised and sorted. The final product was a list of sub-codes which differentiated the master-codes. A sample of the derived coding can be found in Table 4.

4 Interview Results

The interviews and the data analysis were conducted with regard to the initial research questions. This resulted in a coding frame of five master-codes from which three address our research questions directly. In the next subsections, we briefly show the results of the role of security in CP selection, reasons for a moderate interest in security, and the verification of providers' security measures. Since in most of the interviews compliance was strongly connected with security, we also investigated the role of the General Data Protection Regulation (GDPR).

4.1 The Role of Security in CP Selection

The respondents were asked which criteria or requirements they considered when choosing a CP, instead of directly being asked about the role of security. Analogously, the master code "Provider Selection" was extracted from the material with several security related and unrelated sub-codes. The results were selection criteria, of which the ones unrelated to security will only be presented shortly. The most discussed selection criteria were costs (addressed by 5 respondents), size of provider (4) followed by ease of use (3).

Trust: In three interviews the providers' image came up in relation to their trustworthiness, which revealed divided opinions. R1 and R3 provided statements indicating that the image could serve as a proxy for security considerations. *R1: In our region Google did not manage to gain ground, which in my opinion can be contributed to the fact that we are a little bit more sensitive with regard to security and privacy than other countries. So many people shy away when they hear the name "Google" considering them a "data collector".* Similarly, R3 stated that he would consider any large provider except for the Chinese Alibaba cloud. R2 provided the contrary provider's view on this idea. His small company was able to benefit from the image of the local German cloud in the beginning.

Compliance: Non surprisingly, need for security because of compliance appeared referring to regulation authorities, e.g. BaFin or BNetzA (R4, R6, R8).

Availability: Also a great value was placed on the availability of services (R1, R2, R4, R8) in particular over different time zones and with a certain force. Additionally, the statement of R4 even exceeded availability by considering business continuity of the provider to be able to plan for the future.

Confidentiality: The respondents R3 and R4 considered security for the sake of confidentiality of their users' data. *B3: It is about customer data which is located somewhere and one cannot be sure who has access to it. Of course one would like to use cloud services and algorithms to generate an added value out of this data. But on the other hand, one wants to protect the customer from an unauthorized party to gain access to it. I think this is incredibly difficult.* This statement was the only one in the sample expressing a concern for confidentiality apart from any business goals.

Besides selection criteria, several respondents provided insights on how their organisations selected their current CPs. These additionally provided circumstances matter for understanding the provider selection in its context.

Multiple Providers: Among others, it was stressed that current environments consisted of more than one main provider for the sake of independence, availability and freedom of choice (R3, R4). The decision which project or task was done with which provider was a per case decision, depending on the properties of the data and the provider (R4).

Hierarchy: R7 and R5 revealed that the provider decision was made on a higher hierarchical level. Particularly in the case of R7 a provider selection was unnecessary as the company had a strategic partnership with Microsoft.

Convenience: Several respondents admitted that the choice for a CP was partly made by chance, e.g. simply chose a convenient provider to make the first steps in the cloud (R1, R5), because a developer already had some experience (R4) or the company had a voucher (R8). In individual cases these first steps of conveniently testing out a new provider even contradicted corporate requirements and constituted a shadow IT. Despite these tendencies, a security analysis was done retrospectively (R4, R5). Even if it was done retrospectively, the analysis was not only formal but could have changed the decision. *R5: Basically the cloud risk process could have stopped the decision for the product.*

4.2 Reasons for Moderate Interest in Security

The respondents could not be asked why security was only of moderate interest, as security was sooner or later addressed in all the discussions. Nevertheless, most of the answers could be related to "coping with risk". The related topics came up when the respondents were asked about the role of trust or whether they had possible concerns about confidentiality. Most respondents agreed that these concerns do exist but revealed different "coping mechanisms".

Mitigation: Two ways of mitigating the risk raised by respondents were the choice of a large provider and a national or EU-located data centre. In four interviews the location of a data centre came up as a signal of a trustworthy or preferable provider (R2, R3, R4, R5). The assumption, that especially large providers are secure and trustworthy was found in all the interviews except the one with R3. Most respondents argued that large providers invested more in security and thereby also provided a higher level of security than even possible in the own company, which is in line with academic findings [23,38]. Another benefit was stressed by R6 and R8, namely that large companies were also more likely to cover high compensations than small providers in case of a breach.

Responsibility: R2, R5, R7 and R8 agreed that security was not only the responsibility of the CP, but rather a shared one. R2 stressed the differences compared to traditional technologies with regard to responsibility. *R2: Who bears which responsibility often changes in the cloud compared to traditional methods.[...] Before, I either used to run an in-house data centre or I outsourced it.*

R5 stressed the importance of creating awareness in-house for the new technology and its specific risks.

Encryption: Four respondents reported encryption as a mean to secure the cloud. R6 and R8 attached great importance on encrypting their outsourced data and R1 and R2 reported on means of encryption implemented by their clients. Additionally, R2 pointed out the potential drawbacks for the cloud customer. *R2: When we provide the infrastructure only, encryption is mostly in the hands of the customer. But then he has to manage the keys, which represents an additional complexity he has to handle.*

Data Criticality: In addition, some users saw security relatively to the criticality of data they placed into the cloud. R1 and R6 stated that business critical-data was preferably not outsourced at all. *R1: In my opinion, it will always be the case that for a certain part the companies say: "These are my crown jewels, which I don't give away. No matter how much I trust a provider, I want to have these with me".*

Trust: As the opposite side of mitigation, ideas were raised resonating with trust towards the provider. Maybe the most prominent statement to this topic was given by R1: *I believe that many give their providers a few laurels in advance. "Okay they do this on such a large scale and I either I do not trust them per se. In this case I address encryption and other topics. Or as I said, I give them laurels in advance and say, yes this is going to work out"*, assuming that many users trust their providers without any proof. R2, R4, R5 and R8 expressed their belief that the incentives for providers were set in such a way that they cannot afford to make mistakes with customers' data.

Personal Responsibility: R2 tried to explain the popularity of Amazon with the "IBM Effect". *R2: Well I can rely on them (AWS), at least at most times. And when there is a service failure, it applies to everyone and one can say: "Yes, you know it, AWS just had an outage". So it's the IBM effect: "No one ever got fired for buying IBM", applies to AWS nowadays.* R3 agreed with this idea. Finally, independently of mitigation or trust one question had to be included in light of the given answers concerning the importance of security. Throughout some discussions one could have gotten the impression that some companies simply avoided being held accountable in case of a data breach. Therefore the respondents were asked whether there was a personal responsibility or even an intrinsic motivation to pursue security conscientiously. Consequently, the code "Personal Responsibility" was covered with six respondents.

Compliance: The resulting discussions with R1 and R2 were leaned on the fulfillment of GDPR and compliance requirements and both respondents revealed the belief that the choice of a secure provider is rather extrinsically motivated by the need to comply. They also agreed that the regulating authorities still have not drawn any consequences but most likely would do so in the future in order to set an example. *R1: [...] I believe that many (companies) still wait until the first penalties are issued, as surprisingly it (GDPR) did not have that*

many impact yet. [...] I think the first time something happens and jurisdiction is drawn, and a company really has to pay for it, many others will have a second awakening. R4 and R6 agreed that compliance is decisive for the final choice. However, according to R4 intrinsic motivation is individual and depends on the employee's training. *R4: Well it depends on who is dealing with the topic. As I already said, the energy sector has very high security requirements, so if a classic energy economist deals with it, then security and compliance are in his blood. [...] If it is a developer, he may not care. He only asks where to put the data, but does not really think about it himself.* However, R4 adds that in recent years the awareness has risen among all the employees.

4.3 Verification of Providers' Security Measures

The first part of the interviews showed that although security was not the top criterion when selecting a CP, it was present as a requirement. For this reason, it could not be directly asked how the respondents compared different providers with regard to security beforehand, but it could be discussed whether they verified the security levels of their CPs.

Certification: The probably most discussed assurance technique in this sample was certification. According to R1, R2, R4, R6 and R8 two kinds of certification seemed to be of importance when a provider was checked. This was either certification after the ISO norm 27001 or the C5 by BSI (R1, R2, R4, R6), a German governmental agency, which among others incorporates the ISO norm and is combined with an audit. R1 expressed his doubts about C5 being attractive to providers who want to achieve global standardization, as it was a German norm. R4 and R6 agreed that certification in general provided a solid basis for trusting a provider, as for one thing certification institutions could be considered credible and for the other their certification process was very demanding. R2 as well stressed the convenience of certificates but later on also warned of misunderstandings, as one always had to look closely at the coverage. *R2: Another important thing is that certificates are often misunderstood. For instance a 9001 certificate can be done for different domains of my company. I could only certify the administration and in that case a production- or data center is not covered at all.* Moreover, R2's small company could not be certified as the formalization of processes was not possible in the dynamic environment of a start-up. These aspects were also picked up by R8 who criticized exactly that certification was for the most parts focused on processes on paper, which in his view would not provide real security.

Audits: Another assurance technique discussed was external auditing, although it has to be said that the audits most respondents considered were not of technical but rather a financial nature. R6 and R7 for instance stated to have sent public accountants or financial auditors to their providers who apparently only in the broadest sense verified provider security. R1 admitted that he did not know of anyone who really audited their CPs and predicted it rather as a future trend after the clients had made some experiences in the cloud. R7 and R8 stressed the

benefits of a third party audit, namely that an expert was checking the status of a system and giving advice on how to improve it, which was according to R8 an advantage compared to certificates. While R4 doubted the competence of some auditors, R8 pointed out the conflict of interest. *R8: Exactly, it depends on what kind of auditor you get. You can entrust someone who issues an affirmation for you: "Audit accomplished", or you can entrust someone who works conscientiously. The only problem is that the ones who work conscientiously, are often those who are not well received and afterwards have trouble reselling. There is a slight conflict of interest.*

Contracts: It was often discussed in connection to assurance that respondents had contractual agreements with their providers (R4 and R6). R6 added the possibility to contractually seal where data is located and processed. R2 pointed out that contractual agreements were often not only an option but a requirement in light of GDPR, while R4 and R6 gave the important reason for having a contractual agreement, namely that in case of non-fulfillment a compensation was ensured. R1, R2 and R4 mentioned the possibility to contractually include the users' right to visit the data center in person. According to R2 such a clause may be necessary or important to a client, who handles personal data. Nevertheless, the respondents admitted that in reality such a visit hardly ever happened. Additionally, R2 doubted the sense of sending company representatives to visit a data center. *R2: If someone like you or me went there, what would we be supposed to see? If the door is not open somewhere or a cable hanging loosely, we would have no idea how secure this is and whether it is in accordance to the norm.* R1 added that the providers tried to avoid such visits as they considered the interior of their data centre as a company secret. Additionally, checking technical documentation or documentation of processes was found in the interviews (R4, R6, R7).

Tests: Additionally, R4 and R6 talked about security tests as a mean of assurance. *R6: That means that for a cloud service we will not check whether it is externally attackable, as most data centres must have tested this already for about five-, six-, seven-, eight hundred times. What we check is whether the access point we have to the data centre is secure enough.* R4 also stressed that the tests were not done on the CPs' side but on the final application, which was supposed to run in the cloud or as a hybrid application. Both respondents pointed out some drawbacks of penetration-testing, first the costliness and second that such tests could only be run for known cases.

Two respondents stood out with their companies' specific assurance techniques. R5 reported of his companies' own cloud risk process which helped evaluating a provider with regard to the risk he poses to the company and its data. The process incorporated some of the already presented techniques, like demanding a certification and contractually sealing requirements, but more than this, it was a spreadsheet for assessing the likeliness of scenarios and finally presenting the risk imposed by a provider. Finally, the management was in charge of deciding whether this risk was acceptable or not. The other individual measure

was taken by R4's company, which had designed their own questionnaire for CPs comparable to the CAIQ by the CSA.

Finally, besides all the collected assurance techniques it has to be mentioned that several respondents also expressed scepticism when talking about assurance. According to R3 there was no gain from SLAs and contracts, as even if there was a written agreement one had to suffer in case of a data breach in terms of data loss. R4 pointed out the drawback of a third party audit, by telling his own experience with auditors who believed him anything he told them. R7 had doubts about assurance in general and pointed out how the need to control or verify everything although one had outsourced brought unnecessary costliness. Similarly, R8 criticized that certificates do not show real security.

4.4 Compliance and the General Data Protection Regulation

Due to the previous answers, we also elaborate how the GDPR influenced the decisions and to what extent interviewees reported about German and European cloud services which do not transfer data outside of the European Union.

GDPR: According to R2 and R6, a result of the GDPR is that more attention is turned to data protection. R2 claims that the GDPR allows to ensure technical and organisational measures by SLAs more easily.

R1 and R2 agree that since so far data protection authorities have not punished companies by a fine, most companies will assume the first cases will hit large companies and wait for that. R2 was more concerned about written warnings from competitors. R7 reported that his company's data security officer answered to a request about using cloud services that an agreement of the parent company (in Great Britain) with the cloud provider is seen as valid for all subsidiary companies. In contrast, R4 reported that the regulation requires data centres in the EU, which still did not work out for them, because of US employees with access to the stored data. However, they use a CP in Switzerland for non business critical data.

Localisation of CPs: Statements on the localisation of CPs were ambivalent. On the one hand, R3 was concerned about US industrial espionage facilitated by war on terror laws and thus demands a German/European solution with all components (software, hardware) built and run in Germany/EU. This is in line with the report of a "Robin Hood" bonus for a localised offer (R2).

On the other hand R1 and R2 report that at the beginning localisation seemed important, but then lost importance due to data centres in Germany (from the large CPs) and due to observations of other companies seemingly running their cloud services GDPR-compliant with non-EU CPs. An additional argument was that the advantages of localisation can not compensate higher costs (R3, R4, R7), missing features (R1, R2) or development tools (R3) for the German version, customers in the US (R1), and missing trust in the continuity of the service (R4). Many interviewees (R1, R2, R3, R4, R7) were referring to

the "German cloud", a cooperation between Telekom and Microsoft which was ended last year[1].

5 Discussion

Role of Security: With regard to the original question on the role of security in cloud provider selection the collected findings are ambiguous. Selection criteria like usability and costs were expressed straightforwardly and matched the findings of the related work [26,60]. Security however, was never the first answer the respondents extensively engaged in. Neither could they provide concrete security requirements comparable to those found in the related contributions. On the other hand, security as a requirement was present in all the discussions. Moreover, availability and in rare cases confidentiality could be extracted as goals. Two respondents revealed that although security had not been a selection criterion, it was considered in retrospect in some cases, where the companies analysed the services after having tested them first. Moreover, the findings from this sample challenge the idea of a systematic provider selection suggested in related works. In this sample it was rarely the case that providers were compared and evaluated in advance with regard to certain criteria.

Moderate Interest in Security: Some respondents assessed the situation and acted in accordance to the mitigation measures proposed in cloud organizations' technical reports. For instance, one could identify the awareness of the separation of duties and the willingness to employ encryption on the user side. These users were aware that security in the cloud was not only the cloud provider's duty and took own responsibility. On the other hand, namely the capability of a provider to grant compensations speaks however again for a financial interest rather than an intrinsic motivation to establish security. The initial assumption that the requirement on security is extrinsically motivated by compliance was clearly supported by the respondents' answers on personal responsibility. The answers revealed as well a different side to the client provider relationship, which was a great amount of trust towards the cloud provider and the acceptance of risk to a certain extent. The idea that an "IBM effect" exists when choosing Amazon's services indicates that this could be a way for decision makers to be exonerated from responsibility.

Security Assurance: Overall, the respondents revealed to rely on certification, audits, contractual agreements and testing as common means of assurance. Besides those assurance techniques, two respondents presented own company-specific methods. The results from this sample show that except for C5 which is a cloud-specific certificate and audit, the companies rather rely on traditional forms of assurance than cloud-specific ones. Especially contractual agreements are considered a convenient method in order to establish compliance and guarantee for a compensation in case of non-fulfillment. Surprisingly, contractually agreed measures like data center visits are not often undertaken. These findings

[1] https://heise.de/-4152650.

are one more indicator that security and also assurance are overshadowed by compliance, but that at the same time regulation may miss out on establishing real and not only paper-based assurance.

In comparison to the findings from academic literature cloud-specific assurance techniques seemed to have not really thrived in practice. Certification which was most present in the literature review was similarly well accepted among the practitioners as a convenient assurance technique. Testing in terms of application security was also present in both, literature and interviews. However, it is striking but not surprising that neither monitoring nor auditing, which offered many cloud-specific frameworks in literature, were present among the respondents. Contractual agreements could be compared to security SLAs with regard to how they work, except that there are no actual metrics agreed upon but rules.

5.1 Threats to Validity and Limitations

One of the major challenges of conducting the interviews turned out to be finding the right respondents. The ideal respondent given the research questions would have been someone in a C-Level position, who was involved in cloud adoption and knowledgeable about the processes in IT and security. Such persons were difficult to reach or to find time to schedule a face to face interview. In the current sample, respondents from the financial industry are a bit overrepresented and it would have been beneficial to have more respondents from small and medium enterprises. In particular, R8 answered from a perspective of a start-up and could contribute some new ideas. Thus, the interviews should be considered as a first insight and be extended by further interviews with representatives from small- and middle sized companies. Most respondents eventually talked about infrastructure- or platform providers, most likely because in the case of Software-as-a-Service one would rather talk about service- than provider selection.

6 Conclusion

Previous research identified security as a requirement considered by CP customers. Our sample indicates that security may not always be a selection criterion and neither the most decisive one. If considered in the CP selection, then mostly in terms of availability and for the sake of compliance. Especially the focus on compliance it not surprising as it has been observed in other sectors as well [16,54]. Nevertheless, it is certainly a requirement companies have, which manifests itself in cloud use. This is indicated by retrospective analysis and considerations of multiple providers.

CP Selection Process: In our sample we could rarely find any elaborated process of eliciting requirements and then coming to a rational decision which CP to select. Instead, CP were chosen based on vouchers, by chance (just pick on CP for 'testing', but then stick with it), by the management because of established relationships, or because of previous experience from a developer. Even more, some companies make use of many CPs in an unstructured way,

e.g. each department decides by its own. Another pattern we could identify was that companies often try to 'first get into the cloud' and then optimise costs and sometimes security (lift and shift) or try to sort out the collection of different CPs. Further research would be desired to investigate why the methodology proposed by research seems to be rarely used in practise.

For that purpose the different roles in the requirements/decision making process should be investigated in detail and elaborated at which step the relevant methodologies from research were not considered and why.

Assurance: The respondents reported on using more than one assurance technique, combined models from the literature were not present at all. Additionally, they saw flaws in the existing assurance techniques and may not even be acquainted with possible cloud-specific assurance. Thus, the noteworthy finding of this comparison is a divergence between the assurance methods adopted in practice and the cloud-specific ones proposed in literature. It can be speculated whether some academic approaches to assurance have never exceeded their theoretical approach or if they were not able to gain ground in practice yet.

Company Size: Although the results uncover many dimensions and patterns of cloud security, they are not complete. As mentioned earlier, no saturation of interviews could be reached among small and unregulated companies. In contrast, large regulated companies were well represented and most likely contributed to a strong focus on compliance in this analysis. Future work could examine on a larger scale whether and how companies have incorporated security into their provider selection and in particular investigate commonalities and differences between smaller and larger companies.

Big CPs vs. Localisation: It seems that the big CPs are in general trusted by the companies and the idea of a German cloud failed. Companies are trying to setup a compliant way to work with the big CPs. However, one interviewee was concerned about industrial espionage and strongly voted for a European or German CP with all components made in the EU. Further research should unfold the different dimensions of trust, and also investigate to which extent regulations or agreements as the EU–US Privacy Shield influence it.

Gaps Between Research and Practise: In the requirement elicitation and decision making process and in the use of assurance technologies there seems to be a gap between research and practise. This gap is something which seems to be quite common in a lot of areas [52]. Further work should investigate whether this is just a typical finding and already existing ideas can be applied to bridge it [20] or if it is a context specific problem and new ideas are needed.

References

1. Akerlof, G.A.: The market for "lemons": quality uncertainty and the market mechanism. In: Uncertainty in Economics, pp. 235–251. Elsevier (1978)
2. Alhenaki, L., Alwatban, A., Alahmri, B., Alarifi, N.: Security in cloud computing: a survey. Int. J. Comput. Sci. Inf. Secur. (IJCSIS) **17**(4), 67–90 (2019)

3. Anisetti, M., Ardagna, C.A., Damiani, E.: A certification-based trust model for autonomic cloud computing systems. In: 2014 International Conference on Cloud and Autonomic Computing, pp. 212–219 (2014)
4. Anisetti, M., Ardagna, C.A., Damiani, E.: A test-based incremental security certification scheme for cloud-based systems. In: 2015 IEEE International Conference on Services Computing, pp. 736–741 (2015)
5. Anisetti, M., Ardagna, C.A., Damiani, E., Gaudenzi, F., Veca, R.: Toward security and performance certification of open stack. In: 2015 IEEE 8th International Conference on Cloud Computing, pp. 564–571 (2015)
6. Anisetti, M., Ardagna, C.A., Gaudenzi, F., Damiani, E.: A certification framework for cloud-based services. In: Proceedings of the 31st Annual ACM Symposium on Applied Computing, SAC 2016, pp. 440–447. ACM (2016)
7. Ardagna, C.A., Asal, R., Damiani, E., Vu, Q.H.: From security to assurance in the cloud: a survey. ACM Comput. Surv. **48**(1), 2:1–2:50 (2015)
8. Ba, H., Zhou, H., Bai, S., Ren, J., Wang, Z., Ci, L.: jMonAtt: integrity monitoring and attestation of JVM-based applications in cloud computing. In: ICISCE, pp. 419–423 (2017)
9. Bleikertz, S., Mastelic, T., Pape, S., Pieters, W., Dimkov, T.: Defining the cloud battlefield - supporting security assessments by cloud customers. In: IC2E, pp. 78–87 (2013)
10. Briggs, B., Lamar, K., Kark, K., Shaikh, A.: Manifesting legacy: looking beyond the digital era. Technical report, 2018 Global CIO Survey, Deloitte (2018)
11. Casola, V., Benedictis, A.D., Rak, M., Villano, U.: SLA-based secure cloud application development: the SPECS framework. In: SYNASC, pp. 337–344 (2015)
12. CSA: Top threats to cloud computing v1.0. Technical report, Cloud Security Alliance (2010). https://cloudsecurityalliance.org/topthreats/csathreats.v1.0.pdf
13. CSA: The notorious nine: cloud computing top threats in 2013. Technical report, Cloud Security Alliance (2013). https://cloudsecurityalliance.org/download/artifacts/the-notorious-nine-cloud-computing-top-threats-in-2013/
14. CSA: The treacherous 12 - cloud computing top threats in 2016. Technical report, Cloud Security Alliance (2016). https://downloads.cloudsecurityalliance.org/assets/research/top-threats/Treacherous-12_Cloud-Computing_Top-Threats.pdf
15. CSA: Top threats to cloud computing the egregious 11. Technical report, Cloud Security Alliance (2019). https://cloudsecurityalliance.org/download/artifacts/top-threats-to-cloud-computing-egregious-eleven/
16. Dax, J., et al.: IT security status of German energy providers (2017). https://arxiv.org/abs/1709.01254
17. Deng, L., Liu, P., Xu, J., Chen, P., Zeng, Q.: Dancing with wolves: towards practical event-driven VMM monitoring. In: Proceedings of the 13th ACM SIGPLAN/SIGOPS International Conference on VEE, pp. 83–96. ACM (2017)
18. Di Giulio, C., Kamhoua, C., Campbell, R.H., Sprabery, R., Kwiat, K., Bashir, M.N.: IT security and privacy standards in comparison: improving FedRAMP authorization for cloud service providers. In: CCGrid, pp. 1090–1099 (2017)
19. Di Giulio, C., Sprabery, R., Kamhoua, C., Kwiat, K., Campbell, R.H., Bashir, M.N.: Cloud standards in comparison: are new security frameworks improving cloud security? In: CLOUD, pp. 50–57 (2017)
20. Ferguson, J.: Bridging the gap between research and practice. Knowl. Manag. Dev. J. **1**(3), 46–54 (2005)
21. Fernando, R., Ranchal, R., Bhargava, B., Angin, P.: A monitoring approach for policy enforcement in cloud services. In: CLOUD, pp. 600–607 (2017)

22. Ghutugade, K.B., Patil, G.A.: Privacy preserving auditing for shared data in cloud. In: CAST, pp. 300–305 (2016)
23. Gupta, P., Seetharaman, A., Raj, J.R.: The usage and adoption of cloud computing by small and medium businesses. Int. J. Inf. Manag. **33**(5), 861–874 (2013)
24. Haeberlen, T., Dupré, L.: Cloud computing - benefits, risks and recommendations for information security. Technical report, ENISA (2012)
25. Henze, M., et al.: Practical data compliance for cloud storage. In: 2017 IEEE International Conference on Cloud Engineering (IC2E), pp. 252–258 (2017)
26. Hetzenecker, J., Kammerer, S., Amberg, M., Zeiler, V.: Anforderungen an cloud computing Anbieter. In: MKWI (2012)
27. Ismail, U.M., Islam, S., Islam, S.: Towards cloud security monitoring: a case study. In: Cybersecurity and Cyberforensics Conference (CCC), pp. 8–14 (2016)
28. Jakhotia, K., Bhosale, R., Lingam, C.: Novel architecture for enabling proof of retrievability using AES algorithm. In: ICCMC, pp. 388–393 (2017)
29. Jansen, W., Grance, T.: SP 800-144. Guidelines on security and privacy in public cloud computing. Technical report, NIST (2011)
30. Jiang, T., Chen, X., Ma, J.: Public integrity auditing for shared dynamic cloud data with group user revocation. IEEE Trans. Comput. **65**(8), 2363–2373 (2016)
31. Kaaniche, N., Mohamed, M., Laurent, M., Ludwig, H.: Security SLA based monitoring in clouds. In: IEEE EDGE, pp. 90–97 (2017)
32. Kanstrén, T., Lehtonen, S., Savola, R., Kukkohovi, H., Hätönen, K.: Architecture for high confidence cloud security monitoring. In: IC2E, pp. 195–200 (2015)
33. Katopodis, S., Spanoudakis, G., Mahbub, K.: Towards hybrid cloud service certification models. In: IEEE International Conference on Services Computing, pp. 394–399 (2014)
34. Krotsiani, M., Spanoudakis, G.: Continuous certification of non-repudiation in cloud storage services. In: 2014 IEEE 13th International Conference on Trust, Security and Privacy in Computing and Communications, pp. 921–928 (2014)
35. Krutz, R.L., Vines, R.D.: Cloud Security: A Comprehensive Guide to Secure Cloud Computing. Wiley, Hoboken (2010)
36. Kuckartz, U.: Qualitative Inhaltsanalyse: Methoden, Praxis, Computerunterstützung. Beltz Juventa (2016)
37. Kumar, R., Goyal, R.: On cloud security requirements, threats, vulnerabilities and countermeasures: a survey. Comput. Sci. Rev. **33**, 1–48 (2019)
38. Lacity, M.C., Reynolds, P.: Cloud services practices for small and medium-sized enterprises. MIS Q. Exec. **13**(1), 31–44 (2014)
39. Lang, M., Wiesche, M., Krcmar, H.: What are the most important criteria for cloud service provider selection? A Delphi study. In: ECIS (2016)
40. Lee, C., Kavi, K.M., Paul, R.A., Gomathisankaran, M.: Ontology of secure service level agreement. In: 2015 IEEE 16th International Symposium on High Assurance Systems Engineering, pp. 166–172 (2015)
41. Lins, S., Grochol, P., Schneider, S., Sunyaev, A.: Dynamic certification of cloud services: trust, but verify!. IEEE Secur. Priv. **14**(2), 66–71 (2016)
42. Lins, S., Schneider, S., Sunyaev, A.: Trust is good, control is better: creating secure clouds by continuous auditing. IEEE Trans. Cloud Comput. **6**(3), 890–903 (2018)
43. Lins, S., Thiebes, S., Schneider, S., Sunyaev, A.: What is really going on at your cloud service provider? Creating trustworthy certifications by continuous auditing. In: 48th HICSS, pp. 5352–5361 (2015)
44. Luna, J., Suri, N., Iorga, M., Karmel, A.: Leveraging the potential of cloud security service-level agreements through standards. IEEE Cloud Comput. **2**(3), 32–40 (2015)

45. Ma, M., Weber, J., van den Berg, J.: Secure public-auditing cloud storage enabling data dynamics in the standard model. In: DIPDMWC, pp. 170–175 (2016)
46. Mahesh, A., Suresh, N., Gupta, M., Sharman, R.: Cloud risk resilience: investigation of audit practices and technology advances-a technical report. Int. J. Risk Conting. Manag. (IJRCM) **8**(2), 66–92 (2019)
47. Majumdar, S., Madi, T., Wang, Y., Jarraya, Y., Pourzandi, M., Wang, L., Debbabi, M.: User-level runtime security auditing for the cloud. IEEE Trans. Inf. Forensics Secur. **13**(5), 1185–1199 (2018)
48. Meera, G., Geethakumari, G.: A provenance auditing framework for cloud computing systems. In: SPICES, pp. 1–5 (2015)
49. Mohammed, M.M.Z.E., Pathan, A.K.: International center for monitoring cloud computing providers (ICMCCP) for ensuring trusted clouds. In: IEEE 11th International Conference on Ubiquitous Intelligence and Its Associated Workshops, pp. 571–576 (2014)
50. More, S.S., Chaudhari, S.S.: Secure and efficient public auditing scheme for cloud storage. In: CAST, pp. 439–444 (2016)
51. Munoz, A., Mafia, A.: Software and hardware certification techniques in a combined certification model. In: SECRYPT, pp. 1–6 (2014)
52. Norman, D.A.: The research-practice gap: the need for translational developers. Interactions **17**(4), 9–12 (2010)
53. Nugraha, Y., Martin, A.: Towards the classification of confidentiality capabilities in trustworthy service level agreements. In: IC2E, pp. 304–310 (2017)
54. Pape, S., Pipek, V., Rannenberg, K., Schmitz, C., Sekulla, A., Terhaag, F.: Stand zur IT-Sicherheit deutscher Stromnetzbetreiber (2018). http://dokumentix.ub.uni-siegen.de/opus/volltexte/2018/1394/
55. Parasuraman, K., Srinivasababu, P., Angelin, S.R., Devi, T.A.M.: Secured document management through a third party auditor scheme in cloud computing. In: ICECCE, pp. 109–118 (2014)
56. Pasquier, T.F.J., Singh, J., Bacon, J., Eyers, D.: Information flow audit for PaaS clouds. In: IEEE IC2E, pp. 42–51 (2016)
57. Polash, F., Shiva, S.: Building trust in cloud: service certification challenges and approaches. In: 9th International Conference on Complex, Intelligent, and Software Intensive Systems, pp. 187–191 (2015)
58. Ramokapane, K.M., Rashid, A., Such, J.M.: Assured deletion in the cloud: requirements, challenges and future directions. In: CCSW, pp. 97–108. ACM (2016)
59. Rashmi, R.P., Sangve, S.M.: Public auditing system: improved remote data possession checking protocol for secure cloud storage. In: iCATccT, pp. 75–80 (2015)
60. Repschläger, J., Wind, S., Zarnekow, R., Turowski, K.: Developing a cloud provider selection model. In: EMISA (2011)
61. Rewadkar, D.N., Ghatage, S.Y.: Cloud storage system enabling secure privacy preserving third party audit. In: ICCICCT, pp. 695–699 (2014)
62. Rios, E., Mallouli, W., Rak, M., Casola, V., Ortiz, A.M.: SLA-driven monitoring of multi-cloud application components using the MUSA framework. In: IEEE 36th ICDCSW, pp. 55–60 (2016)
63. Rizvi, S.S., Bolish, T.A., Pfeffer III, J.R.: Security evaluation of cloud service providers using third party auditors. In: Second International Conference on Internet of Things, Data and Cloud Computing, pp. 106:1–106:6 (2017)
64. Ryoo, J., Rizvi, S., Aiken, W., Kissell, J.: Cloud security auditing: challenges and emerging approaches. IEEE Secur. Priv. **12**(6), 68–74 (2014)

65. Schneider, S., Lansing, J., Gao, F., Sunyaev, A.: A taxonomic perspective on certification schemes: development of a taxonomy for cloud service certification criteria. In: HICSS, pp. 4998–5007 (2014)
66. Sen, A., Madria, S.: Data analysis of cloud security alliance's security, trust & assurance registry. In: ICDCN, pp. 42:1–42:10. ACM (2018)
67. Sotiriadis, S., Lehmets, A., Petrakis, E.G.M., Bessis, N.: Unit and integration testing of modular cloud services. In: AINA, pp. 1116–1123 (2017)
68. Stephanow, P., Khajehmoogahi, K.: Towards continuous security certification of software-as-a-service applications using web application testing techniques. In: AINA, pp. 931–938 (2017)
69. Thendral, G., Valliyammai, C.: Dynamic auditing and updating services in cloud storage. In: International Conference on Recent Trends in Information Technology, pp. 1–6 (2014)
70. Tung, Y., Lin, C., Shan, H.: Test as a service: a framework for web security TaaS service in cloud environment. In: 2014 IEEE 8th International Symposium on Service Oriented System Engineering, pp. 212–217 (2014)
71. Zhang, H., Manzoor, S., Suri, N.: Monitoring path discovery for supporting indirect monitoring of cloud services. In: IEEE IC2E, pp. 274–277 (2018)
72. Zhang, H., Trapero, R., Luna, J., Suri, N.: deQAM: a dependency based indirect monitoring approach for cloud services. In: IEEE SCC, pp. 27–34 (2017)

B.8 Selecting a Secure Cloud Provider: An Empirical Study and Multi Criteria Approach

Sebastian Pape, Federica Paci, Jan Juerjens, and Fabio Massacci. Selecting a secure cloud provider: An empirical study and multi criteria approach. *Information*, 11(5), 05 2020. doi: 10.3390/info11050261. URL https://www.mdpi.com/2078-2489/11/5/261. Section Information Applications, Special Issue Cloud Security Risk Management

information

MDPI

Article

Selecting a Secure Cloud Provider—An Empirical Study and Multi Criteria Approach

Sebastian Pape [1,*], Federica Paci [2], Jan Jürjens [3] and Fabio Massacci [4]

1 Faculty of Economics and Business, Goethe University Frankfurt, 60323 Frankfurt, Germany
2 Department of Computer Science, University of Verona, 37134 Verona, Italy; federicamariafrancesca.paci@univr.it
3 Faculty of Computer Science, University of Koblenz, 56070 Koblenz, Germany & Fraunhofer ISST, 44227 Dortmund, Germany; juerjens@uni-koblenz.de
4 Department of Information Sciences and Engineering, University of Trento, 38123 Trento, Italy; fabio.massacci@unitn.it
* Correspondence: sebastian.pape@m-chair.de; Tel.: +49-69-798-34668

Received: 1 April 2020; Accepted: 6 May 2020; Published: 11 May 2020

check for updates

Abstract: Security has become one of the primary factors that cloud customers consider when they select a cloud provider for migrating their data and applications into the Cloud. To this end, the Cloud Security Alliance (CSA) has provided the Consensus Assessment Questionnaire (CAIQ), which consists of a set of questions that providers should answer to document which security controls their cloud offerings support. In this paper, we adopted an empirical approach to investigate whether the CAIQ facilitates the comparison and ranking of the security offered by competitive cloud providers. We conducted an empirical study to investigate if comparing and ranking the security posture of a cloud provider based on CAIQ's answers is feasible in practice. Since the study revealed that manually comparing and ranking cloud providers based on the CAIQ is too time-consuming, we designed an approach that semi-automates the selection of cloud providers based on CAIQ. The approach uses the providers' answers to the CAIQ to assign a value to the different security capabilities of cloud providers. Tenants have to prioritize their security requirements. With that input, our approach uses an Analytical Hierarchy Process (AHP) to rank the providers' security based on their capabilities and the tenants' requirements. Our implementation shows that this approach is computationally feasible and once the providers' answers to the CAIQ are assessed, they can be used for multiple CSP selections. To the best of our knowledge this is the first approach for cloud provider selection that provides a way to assess the security posture of a cloud provider in practice.

Keywords: cloud service provider; security self-assessment; security assessment; risk assessment

1. Introduction

Cloud computing has become an attractive paradigm for organisations because it enables "*convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort* [1]". However, security concerns related to the outsourcing of data and applications to the cloud have slowed down cloud adoption. In fact, cloud customers are afraid of loosing control over their data and applications and of being exposed to data loss, data compliance and privacy risks. Therefore, when it comes to select a cloud service provider (CSP), cloud customers evaluate CSPs first on security (82%), and data privacy (81%) and then on cost (78%) [2]. This means that a cloud customer will more likely engage with a CSP that shows the best capabilities to fully protect information assets in its cloud service offerings. To identify the "ideal" CSP, a customer has first to assess and compare

the security posture of the CSPs offering similar services. Then, the customer has to select among the candidate CSPs, the one that best meets his security requirements.

Selecting the most secure CSP is not straightforward. When the tenant outsources his services to a CSP, he also delegates to the CSP the implementation of security controls to protect his services. However, since the CSP's main objective is to make profit, it can be assumed that he does not want to invest more than necessary in security. Thus, there is a tension between tenant and CSP on the provision of security. In addition, for security compared to other providers' attributes like cost or performance there are no measurable and precise metrics to quantify it [3]. The consequences are twofold. It is not only hard for the tenant to assess the security of outsourced services, it is also hard for the CSP to demonstrate his security capabilities and thus to negotiate a contract. Thus, even if a CSP puts a lot of effort in security, it will be hard for him to demonstrate it, since malicious CSPs will pretend to do the same. This imbalance of knowledge is known as information asymmetry [4] and together with the cost of cognition to identify a good provider and negotiate a contract [5] has been widely studied in economics.

Furthermore, information gathering on the security of a provider is not easy because there is no standard framework to assess which security controls are supported by a CSP. The usual strategy for the cloud customer is to ask the CSP to answer a set of questions from a proprietary questionnaire and then try to fix the most relevant issues in the service level agreements. But this makes the evaluation process inefficient and costly for the customers and the CSPs.

In this context, the Cloud Security Alliance (CSA) has provided a solution to the assessment of the security posture of CSPs. The CSA published the Consensus Assessments Initiative Questionnaire (CAIQ), which consists of questions that providers should answer to document which security controls exist in their cloud offerings. The answers of CSPs to CAIQ could be used by tenants for selecting the provider the best suit their security needs.

However, there are many CSPs offering the same service—Spamina Inc. lists around 850 CSPs worldwide. While it can be considered acceptable to manually assess and compare the security posture of an handful of providers, this task becomes unfeasible when the number of providers grows up to hundreds. As a consequence, many tenants do not have an elaborated process to select a secure CSP based on security requirement elicitation. Instead, often CSPs are chosen by chance or the tenant just sticks to big CSPs [6]. Therefore, there is the need for an approach that helps cloud customers in comparing and ranking CSPs based on the level of security they offer.

The existing approaches to CSP ranking and selection either do not consider security as a relevant criteria for selection or they do but do not provide a way to assess security in practice. To the best of our knowledge there are no approaches that have used CAIQs to assess and compare the security capabilities of CSPs.

Hence, we investigate in this paper whether manually comparing and ranking CSPs based on CAIQ's answers is feasible in practice. For this aim we have conducted an empirical study that has shown that manually comparing CSPs based on CAIQ is too time consuming. To facilitate the use of CAIQ to compare and ranking CSPs, we have proposed an approach that automates the processing of CAIQ's answers. The approach uses CAIQ's answers to assign a value to the different security capabilities of CSPs and then uses an Analytic Hierarchy Process (AHP) to compare and rank the providers based on those capabilities.

The contribution of this paper is threefold. First, we discuss the issues related to processing CAIQ for provider selection that could hinder its adoption in practice. Second, we refined the security categories used to classify the questions in the CAIQ into a set of categories that can be directly mapped to low-level security requirements. Then, we propose an approach to CSP comparing and ranking that assigns a weight to the security categories based on CAIQ's answers.

To the best of our knowledge, our approach is the only one which provides an effective way to measure the level of security of a provider.

The rest of the paper is structured as follows. Section 2 presents related work and Section 3 discusses the issues related to processing CAIQs. Then, Section 4 presents the design and the results of the experiment and discusses the implications that our results have for security-aware provider selection. Section 5 introduces our approach to comparing and ranking CSPs' security. We evaluate it in Sections 6 and 7 concludes the paper and outlines future works.

In the in Appendix A we give an illustrative example for the application of our approach.

2. Related Work

The problem of service selection has been widely investigated both in the context of web services and cloud computing. Most of the works based the selection on Quality of Service (QoS) but adopt different techniques to comparing and ranking CSPs such as genetic algorithms [7], ontology mapping [8,9], game theory [10] and multi-criteria decision making [11]. In contrast, only few works considered security as a relevant criteria for the comparison and ranking of CSPs [12–18] but none of them provided a way to assess and measure the security of a CSP in practice.

Sundareswaran et al. [12] proposed an approach to select an optimal CSP based on different features including price, QoS, operating systems and security. In order to select the best CSP they encode the property of the providers and the requirements of the tenant as bit array. Then to identify the candidate providers, they find the service providers whose properties encoding are the k-nearest neighbours of the encoding of the tenant's requirements. However, Sundareswaran et al., do not describe how an overall score for security is computed, while in our approach overall security level of a CSP is computed based on the security controls that the provider declares to support in the CAIQ.

More recently, Ghosh et al. [13] proposed SelCSP, a framework that supports cloud customers in selecting the provider that minimises the security risk related to the outsourcing of their data and application to the CSP. The approach consists in estimating the interaction risk the customer is exposed to if it decides to interact with a CSP. The interaction is computed based on the trustworthiness the customer places in the provider and the competence of the CSP. The trustworthiness is computed based on direct and indirect ratings obtained through either direct interaction or other customers' feedback. The competence of the CSP is estimated from the transparency of SLAs. The CSP with minimum interaction risk is the one ideal for the cloud customer. Similarly to us, to estimate confidence Ghosh et al., have identified a set of security categories and mapped those categories to low-level security controls supported by the CSPs. However, they do not mention how a value can be assigned to the security categories based on the security controls. Mouratidis et al. [19] describe a framework to select a CSP based on security and privacy requirements. They provide a modelling language and a structured process, but only give a vague description how a structured security elicitation at the CSP works. Akinrolabu [20] develops a framework for supply-chain risk assessment which can also be used to assess the security of different CSPs. For each CSP a score has to be determined for nine different dimensions. However, they do not mention how a value can be assigned to each security dimension. Habib et al. [18] also propose an approach to compute a trustworthiness score for CSPs in terms of different attributes, for example, compliance, data governance, information security. Similarly to us, Habib et al. use CAIQ as a source to assign a value to the attributes on the basis of which the trustworthiness is computed. However, in their approach the attributes match the security domains in the CAIQ and therefore a tenant has to specify its security requirements in terms of the CAIQ security domains. In our approach, we do not have such a limitation: the tenant specifies his security requirements that are then mapped to security categories, that can be mapped to specific security features offered by a CSP. Mahesh et al. [21] investigate audit practices, map the risk to technology that mitigates the risk and come up with a list of efficient security solutions. However, their approach is used to compare different security measures and not different CSPs. Bleikertz et al. [22] support cloud customers with the security assessments. Their approach is focused on a systematic analysis of attacks and parties in cloud computing to provide a better understanding of attacks and find new ones.

Other approaches [14–16] focus on identifying a hierarchy of relevant attributes to compare CSPs and then use multi-criteria decision making techniques to rank them based on those attributes.

Costa et al. [14] proposed a multi-criteria decision model to evaluate cloud services based on the MACBETH method. The services are compared with respect to 19 criteria including also some aspects of security like data confidentiality, data loss and data integrity. However, the MACBETH approach does not support the automatic selection of the CSP because it requires the tenant to give for each evaluation criteria a neutral reference level and a good reference level and to rate the attractiveness of each criteria. While in our approach the input provided by the tenant is minimised: the tenant only specifies the security requirements and their importance and then our approach automatically compares and ranks the candidate CSPs.

Garg et al. proposed a selection approach based on the Service Measurement Index (SMI) [23,24] developed by the Cloud Services Measurement Initiative Consortium (CSMIC) [25]. SMI aims to provide a standard method to measure cloud-based business services based on an organisation's specific business and technology requirements. It is a hierarchical framework consisting of seven categories which are refined into a set of measurable key performance indicators (KPI). Each KPI gets a score and each layer of the hierarchy gets weights assigned. The SMI is then calculated by multiplying the resulting scores by the assigned weights. Garg et al. have extended the SMI approach to derive the relative service importance values from KPIs, and then use the Analytic Hierarchy Process (AHP) [26,27] for ranking the services. Furthermore, they have distinguished between essential, where KPI values are required, and non-essential attributes. They have also explained how to handle the lack of KPI values for non-essential attributes. Built upon this approach, Patiniotakis et al. [16] discuss an alternative classification based on the fuzzy AHP method [28,29] to handle fuzzy KPIs' values and requirements. To assess security and privacy, Patiniotakis et al. assume that a subset of the controls of the cloud control matrix is referenced as KPIs and that the tenant should ask the provider (or get its responses from the CSA STAR registry) and assign each answer a score and a weight.

As the approaches to CSP selection proposed in References [15–17], our approach adopts a multi-criteria decision model based on AHP to rank the CSPs. However, there are significant differences. First, we refine the categories proposed to classify the questions in the CAIQ into sub-categories that represent well-defined security aspects like access control, encryption, identity management, and malware protection that have been defined by security experts. Second, a score and weight to these categories is automatically assigned based on the answers that providers give to corresponding questions in the CAIQ. This reduces the effort for the cloud customer who can rely on the data published in CSA STAR rather than interviewing the providers to assess their security posture.

Table 1 provides and overview of the mentioned related work. The columns "dimension" list if the approach considers security and/or other dimensions, the column "data security" lists if the approach proposes a specific method how to evaluate security and the column "security categories" lists how many different categories are considered for security.

In summary, to the best of our knowledge, our approach is the first approach to CSP selection that provides an effective way to measure the security of a provider. Our approach could be used as a building block for the existing approaches to CSP selection that consider also other providers' attributes like cost and performance.

Table 1. Comparison of Different cloud service provider (CSP) Comparison/Selection Approaches.

		Dimensions		Security	
Reference	**Method**	**Other**	**Security**	**Data**	**Categories**
Anastasi et al. [7]	genetic algorithms	✓	✗	✗	✗
Ngan and Kanagasabai [8]	ontology mapping	✓	✗	✗	✗
Sim [9]	ontology mapping	✓	✗	✗	✗
Wang and Du [10]	game theory	✓	✗	✗	✗
Karim et al. [11]	MCDM [1]	✓	✗	✗	✗
Sundareswaran et al. [12]	k-nearest neighbours	✓	✓	✗	✗
Ghosh et al. [13]	minimize interaction risk	✓	✓	✗	12
Costa et al. [14]	MCDM [1]	✓	✓	✗	3
Garg et al. [15]	MCDM [1]	✓	✓	✗	7
Patiniotakis et al. [16]	MCDM [1]	✓	✓	✗	1
Wittern et al. [17]	MCDM [1]	✓	✓	✗	unspec.
Habib et al. [18]	trust computation	✓	✓	(✓) [2]	11
Mouratidis et al. [19]	based on Secure Tropos	✗	✓	✗	unspec.
Akinrolabu et al. [20]	risk assessment	✗	✓	✗	9
Our Approach	MCDM [1]	✗	✓	✓	flexible

[1] multi-criteria decision making. [2] Data source (CAIQ) specified, but only yes/no considered and no specific algorithm specified.

3. Standards and Methods

In the first subsection we introduce the Cloud Security Alliance (CSA), the Cloud Controls Matrix (CCM) and the Consensus Assessments Initiative Questionnaire (CAIQ). In the second subsection, we discuss the issues related to the use of CAIQs to compare and ranking CSPs' security.

3.1. Cloud Security Alliance's Consensus Assessments Initiative Questionnaire

The Cloud Security Alliance is a non-profit organisation with the aim to promote best practices for providing security assurance within Cloud Computing [30]. To this end, the Cloud Security Alliance has provided the Cloud Controls Matrix [31] and the Consensus Assessments Initiative Questionnaire [32]. The CCM is designed to guide cloud vendors in improving and documenting the security of their services and to assist potential customers in assessing the security risks of a CSP.

Each control consists of a control specification which describes a best practice to improve the security of the offered service. The controls are mapped to other industry-accepted security standards, regulations, and controls frameworks, for example, ISO/IEC 27001/27002/27017/27018, NIST SP 800-53, PCI DSS, and ISACA COBIT.

Controls covered by the CCM are preventive, to avoid the occurrence of an incident, detective, to notice an incident and corrective, to limit the damage caused by the incident. Controls are in the ranges of legal controls (e.g., policies), physical controls (e.g., physical access controls), procedural controls (e.g., training of staff), and technical controls (e.g., use of encryption or firewalls).

For each control in the CCM the CAIQ contains an associated question which is in general a 'yes or no' question asking if the CSP has implemented the respective control. Figure 1 shows some examples of questions and answers. Tenants may use this information to assess the security of CSPs whom they are considering contracting.

As of today, there are two relevant versions of the CAIQ: version 1.1 from December 2010 and version 3.0.1 from July 2014. CAIQ version 1.1 consists of 197 questions in 11 domains (see Table 2), while CAIQ version 3.0.1 instead consists of 295 questions grouped in 16 domains (see Table 3). In November 2019 version 3.1 of the CAIQ was published and it was stated that 49 new questions were added, and 25 existing ones were revised. Furthermore, with CAIQ-Lite, there exists a smaller version consisting of 73 Questions covering the same 16 Control Domains.

CID	Consensus Assessment Questions	Response	Comments and Notes
CO-01.1	Do you produce audit assertions using a structured, industry accepted format (ex. CloudAudit/A6 URI Ontology, CloudTrust, SCAP/CYBEX, GRC XML, ISACA's Cloud Computing Management Audit/Assurance Program, etc.)?	Yes	Independent internal and external audits are scheduled and conducted with audit assertions produced following ISACA's Cloud Computing Management Audit/Assurance Program and ISO 27001.
CO-02.1	Do you allow tenants to view your SAS70 Type II/SSAE 16 SOC2/ISAE3402 or similar third party audit reports?	Yes	Independent audit reports produced by external security consultant and certification body are available for viewing by tenants upon request.
CO-02.2	Do you conduct network penetration tests of your cloud service infrastructure regularly as prescribed by industry best practices and guidance?	Yes	Penetration Testing and Vulnerability Assessment on the cloud service infrastructure at network, operating systems and application levels are conducted half-yearly by independent security consultant.
CO-02.3	Do you conduct regular application penetration tests of your cloud infrastructure as prescribed by industry best practices and guidance?	Yes	Web Application Penetration Testing and Vulnerability Assessment on the cloud service infrastructure is conducted on a half-year basis by independent security consultant.
CO-02.4	Do you conduct internal audits regularly as prescribed by industry best practices and guidance?	Yes	Internal audits are conducted at least annually using the ISACA Cloud Computing Management Audit / Assurance Program and ISO 27001 as the basis of evaluation.
CO-02.5	Do you conduct external audits regularly as prescribed by industry best practices and guidance?	Yes	External audits are conducted at least annually using ISO 27001 as the basis of evaluation.
CO-02.6	Are the results of the network penetration tests available to tenants at their request?	Yes	Penetration Testing and Vulnerability Assessment Reports are available for viewing by tenants upon request.
CO-02.7	Are the results of internal and external audits available to tenants at their request?	Yes	Internal and external audit reports are available for viewing by tenants upon request.
CO-03.1	Do you permit tenants to perform independent vulnerability assessments?	Yes	Tenants can perform independent vulnerability assessments of their own virtual infrastructure or equipment.
CO-03.2	Do you have external third-party conduct vulnerability scans and periodic penetration tests on your applications and networks?	Yes	Penetration testing and vulnerability assessment on the applications and networks of the cloud computing infrastructure are conducted by external third party security consultant at least annually.
CO-04.1	Do you maintain liaisons and points of contact with local authorities in accordance with contracts and appropriate regulations?	Yes	Contact list of local authorities is maintained and updated regularly.

(**a**) Snapshot of a CAIQ version 1.1

Control Domain	Control ID	Question ID	Control Specification	Consensus Assessment Questions	Consensus Assessment Answers		
					Yes	No	Not Applicable
Application & Interface Security *Application Security*	AIS-01	AIS-01.1	Applications and programming interfaces (APIs) shall be designed, developed, deployed, and tested in accordance with leading industry standards (e.g., OWASP for web applications) and adhere to applicable legal, statutory, or regulatory compliance obligations.	Do you use industry standards (i.e. OWASP Software Assurance Maturity Model, ISO 27034) to build in security for your Systems/Software Development Lifecycle (SDLC)?			
		AIS-01.2		Do you use an automated source code analysis tool to detect security defects in code prior to production?			
		AIS-01.3		Do you use manual source-code analysis to detect security defects in code prior to production?			
		AIS-01.4		Do you verify that all of your software suppliers adhere to industry standards for Systems/Software Development Lifecycle (SDLC) security?			
		AIS-01.5		(SaaS only) Do you review your applications for security vulnerabilities and address any issues prior to deployment to production?			
Application & Interface Security	AIS-02	AIS-02.1	Prior to granting customers access to data, assets, and information systems, identified security, contractual, and regulatory requirements for customer access shall be	Are all identified security, contractual, and regulatory requirements for customer access contractually addressed and remediated prior to granting customers access to data, assets, and information systems?			
		AIS- 02.2		Are all requirements and trust levels for customers' access defined and documented?			
Application & Interface Security *Data Integrity*	AIS-03	AIS-03.1	Data input and output integrity routines (i.e., reconciliation and edit checks) shall be implemented for application interfaces and databases to prevent manual or systematic processing errors, corruption of data, or misuse.	Does your data management policies and procedures require audits to verify data input and output integrity routines?			
		AIS-03.2		Are data input and output integrity routines (i.e. MD5/SHA checksums) implemented for application interfaces and databases to prevent manual or systematic processing errors or corruption of data?			
Application & Interface Security *Data Security / Integrity*	AIS-04	AIS-04.1	Policies and procedures shall be established and maintained in support of data security to include (confidentiality, integrity, and availability) across multiple system interfaces, jurisdictions, and business functions to prevent improper disclosure, alteration, or destruction.	Is your Data Security Architecture designed using an industry standard (e.g., CDSA, MULITSAFE, CSA Trusted Cloud Architectural Standard, FedRAMP, CAESARS)?			
Audit Assurance & Compliance *Audit Planning*	AAC-01	AAC-01.1	Audit plans shall be developed and maintained to address business process disruptions. Auditing plans shall focus on reviewing the effectiveness of the implementation of security operations. All audit activities must be agreed upon prior to executing any audits.	Do you develop and maintain an agreed upon audit plan (e.g., scope, objective, frequency, resources,etc.) for reviewing the efficiency and effectiveness of implemented security controls?			
		AAC-01.2		Does your audit program take into account effectiveness of implementation of security operations?			
Audit Assurance & Compliance *Independent Audits*	AAC-02	AAC-02.1	Independent reviews and assessments shall be performed at least annually to ensure that the organization addresses nonconformities of established policies, standards, procedures, and compliance obligations.	Do you allow tenants to view your SOC2/ISO 27001 or similar third-party audit or certification reports?			
		AAC-02.2		Do you conduct network penetration tests of your cloud service infrastructure at least annually?			
		AAC-02.3		Do you conduct application penetration tests of your cloud infrastructure regularly as prescribed by industry best practices and guidance?			
		AAC-02.4		Do you conduct internal audits at least annually?			
		AAC-02.5		Do you conduct independent audits at least annually?			
		AAC-02.6		Are the results of the penetration tests available to tenants at their request?			
		AAC-02.7		Are the results of internal and external audits available to tenants at their request?			

(**b**) Snapshot of a CAIQ version 3.1

Figure 1. Consensus Assessments Initiative Questionnaire (CAIQ) questionnaires.

Table 2. Cloud Controls Matrix (CCM)-Item and CAIQ-Question Numbers per Domain (version 1.1).

ID	Domain	CCM-Items	CAIQ-Questions
CO	Compliance	6	16
DG	Data Governance	8	16
FS	Facility Security	8	9
HR	Human Resources	3	4
IS	Information Security	34	75
LG	Legal	2	4
OP	Operations Management	4	9
RI	Risk Management	5	14
RM	Release Management	5	6
RS	Resiliency	8	12
SA	Security Architecture	15	32
	Total	98	197

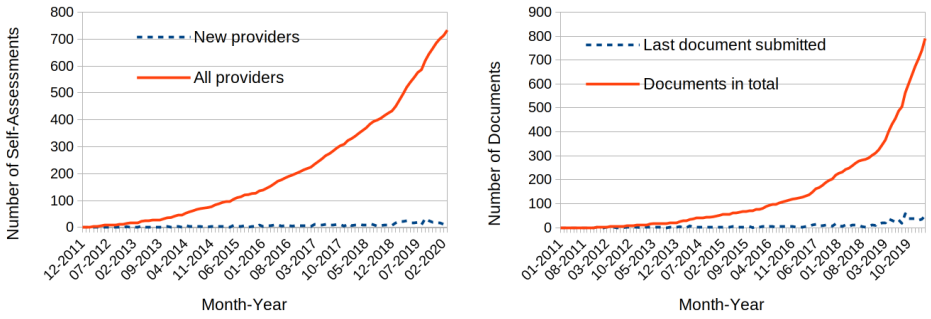
Table 3. Cloud Controls Matrix (CCM)-Item and CAIQ-Question Numbers per Domain (version 3.1).

ID	Domain	CCM	CAIQ
AIS	Application & Interface Security	4	9
AAC	Audit Assurance & Compliance	3	13
BCR	Business Continuity Management & Operational Resilience	11	22
CCC	Change Control & Configuration Management	5	10
DSI	Change Control & Configuration Management	7	17
DCS	Datacenter Security	9	11
EKM	Encryption & Key Management	4	14
GRM	Governance and Risk Management	11	22
HRS	Human Resources	11	24
IAM	Identity & Access Management	13	40
IVS	Infrastructure & Virtualization Security	13	33
IPY	Interoperability & Portability	5	8
MOS	Mobile Security	20	29
SEF	Security Incident Management, E-Discovery & Cloud Forensics	5	13
STA	Supply Chain Management, Transparency and Accountability	9	20
TVM	Threat and Vulnerability Management	3	10
	Total	133	295

CAIQ version 3.0.1 contains a high level mapping to CAIQ version 1.1, but there is no direct mapping of the questions. Therefore, we mapped the questions. In order to determine the differences, we computed the Levenshtein distance (The Levenshtein distance is a string metric which measures the difference between two strings by the minimum number of single-character edits (insertions, deletions or substitutions) required to change one string into the other) [33] between each question from version 3.0.1 and version 1.1. The analysis shows that out of the 197 questions of CAIQ version 1.1 one question was a duplicate, 15 were removed, 12 were reformulated, 79 have undergone editorial changes (mostly Levenshtein distance less than 25), and 90 were taken over unchanged. Additionally 114 new questions were introduced to CAIQ version 3.0.1.

The CSA provides a registry, the Cloud Security Alliance Security, Trust and Assurance Registry (STAR), where the answers to the CAIQ of each participating provider are listed. As shown in Figure 2, the STAR is continuously updated. The overview of answers to CAIQ submitted to STAR in Figure 2 shows that from the beginning in 2011 each year there are more providers contributing to it. At the beginning of October 2014 there were 85 documents in STAR: 65 answers to CAIQ, 10 statements to the CCM, and 10 STAR certifications, where the companies did not publish corresponding self-assessments. In March 2020, there were 733 providers listed with 690 CAIQs (53 versions 1.* or 2515 version 3.0.1,

122 version 3.1), and 106 certifications/attestations. Some companies list the self-assessment along with their certification, some do not provide their self-assessment when they got a certification.



(**a**) Providers (**b**) Documents

Figure 2. Submissions to Security, Trust and Assurance Registry (STAR).

3.2. Processing the CAIQ

Each CAIQ is stored in a separate file with a unique URL. Thus, there is no way to get all CAIQs in a bunch and no single file containing all the answers. Therefore, we had to manually download the CAIQs with some tool support. After downloading, we extracted the answers to the questions and stored them in an SQL database. A small number of answers was not in English and we disregarded them when evaluating the answers.

One challenge was, that there was no standardization of the document format. In October 2014, the 65 answers to CAIQ were in various document formats (52 XLS, 7 PDF, 5 XLS+PDF, 1 DOC). In March 2020, the majority of the document formats was based on Microsoft Excel (615), but there were also others (41 PDFs, 33 Libre Office documents (33), 1 DOC). Besides the different versions, that is, version 1.1 and version 3.0.1, another issue was that many CSP do not comply with the standard format for the answers proposed by the CSA. This makes it not trivial to determine whether a CSP implements a given security control.

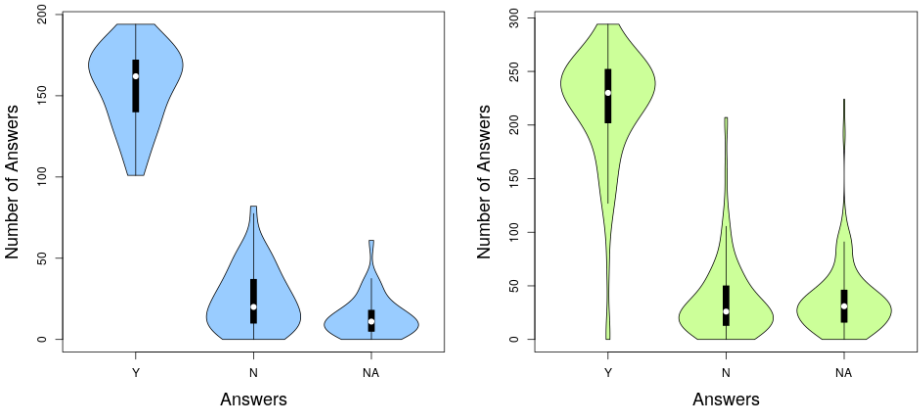
For CAIQ version 1.1 the CSA intended the CSPs to use one column for yes/no/not applicable (Y/N/NA) answers and one column for additional, optional comments (C) when answering the CAIQ. But only a minority (17 providers) used it that way. The majority (44 providers) used only a single column which mostly (22 providers), partly (11 providers) or not at all (11 providers) included an explicit Y/N/NA answer. For CAIQ version 3.0.1 the CSA has introduced a new style: three columns where the provider should indicate whether yes, no or not applicable holds, followed by a column for optional comments. So far, this format for answers seems to work better, since most providers answering CAIQ version 3.0.1 followed it, however, since some providers merged cells, added or deleted columns or put their answer in other places, the answers to the CAIQ can not be gather automatically.

To make it even harder for a customer to determine whether a CSP supports a given security control, the providers did not follow a unique scheme for answers. For example to questions of the kind "Do you provide [some kind of documentation] to the tenant?" some provider answered "Yes, upon request" when others answered "No, only on request". Similarly, some questions asking if controls are in place were answered by some providers with "Yes, starting from [Date in the future]" while others answered "No, not yet". However, these are basically the same answers, but expressed differently. Similar issues could be found for various other questions, too.

Additionally, some providers did not provide a clear answer. For example, some providers claim that they have to clarify some questions with a third party or did not provide answers for questions

at all. Some providers also make use of Amazon AWS (e.g., Acquia, Clari, Okta, Red Hat, Shibumi) but gave different answers when referring to controls implemented by Amazon as IaaS-Provider or did not give an answer and just referred to Amazon.

In order to facilitate the CSPs' answers for comparison and ranking, we give a brief overview of the processed data. Figure 3 (cf. Section 5.4 for information how we processed the data) shows the distribution of the CSPs' answers to the CAIQ. Neglecting the number of questions, there is no huge difference between the distribution in the different versions of the questionnaires. The majority of controls seem to be in place, since "yes" is the most common answer. It can also be seen that the deviation of all answers is quite large which suits to the observation that they are not equally distributed. Regarding the comments on average every second answer had a comment. However, we noticed that comments are a double edge sword: sometimes they help to clarify an answer because they provide the rationale for the answer while at other times they make the answer unclear because they provide information that is conflicting with the yes/no-answer.

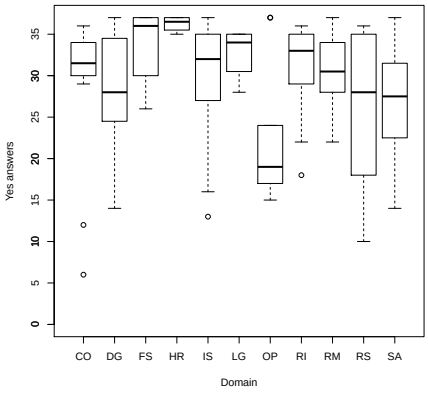


(**a**) CAIQ 1.1, n = 37 (**b**) CAIQ 3.0.1, n = 189

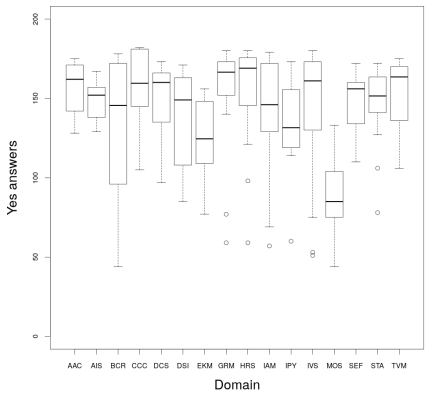
Figure 3. Distribution of Answers per Provider of the CAIQ as Violin-/Boxplot.

We also grouped questions by their domain (x-axis) and for each question within that domain determined the number of providers (y-axis) who answered with yes, no or not applicable. The number of questions per domain can be seen in Tables 2 and 3. Figure 4 shows that for most domains, questions with mostly yes answers dominate (e.g., the domain "human resources" (HR) contains questions with 35 to 37 yes answers from a total of 37 providers (cf. Figure 4a). The domain of "operation management" (OP) holds questions with a significant lower count of yes answers due to questions with many NA answers (cf. Figure 4e), similarly to the domain of "mobile security" (MOS) in version 3.0.1 (cf. Figure 4f). The domains "data governance" (DG), "information security" (IS), "resilience" (RS) and "security architecture" (SA) share a larger variance that means that they contain questions with mostly yes answers as well as questions with only some yes answers.

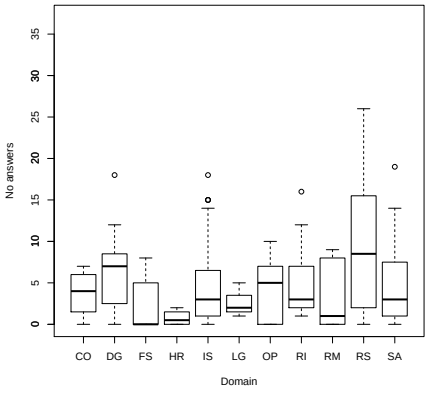
The above issues indicate that gathering information on the CSPs' controls and especially comparing and ranking the security of CSPs using the answers to CAIQ is not straight forward. For this reason, we have conducted a controlled experiment to assess whether it is feasible in practice to select a CSP using CAIQ. We also tested if comments help to determine if a security control is supported or not by CPSs.



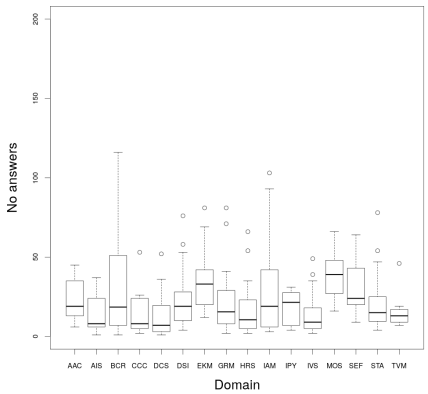
(**a**) Yes Answers, CAIQ v1.1, n = 37



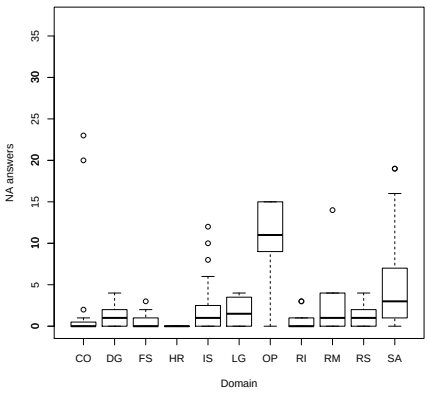
(**b**) Yes Answers, CAIQ v3.0.1, n = 189



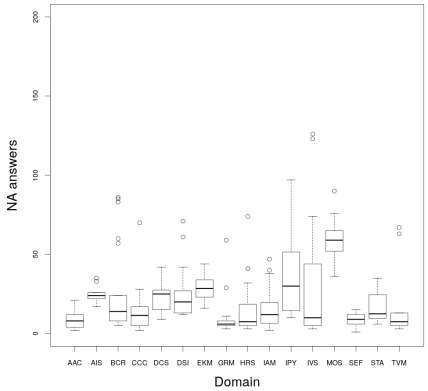
(**c**) No Answers, CAIQ v1.1, n = 37



(**d**) No Answers, CAIQ v3.0.1, n = 189



(**e**) NA Answers, CAIQ v1.1, n = 37



(**f**) NA Answers, CAIQ v3.0.1, n = 189

Figure 4. Distribution of Answers per Question grouped by Domain of CAIQ v1.1 and v3.0.1.

4. Empirical Study on Cloud Service Provider Selection

In this section we report on an empirical study conducted to evaluate the actual and perceived effectiveness of the CSP selection process based on the CAIQ. The perceived effectiveness of the selection process is assessed in terms of perceived ease of use and perceived usefulness.

4.1. Research Questions

The main research questions that we want to address in our study are:

- RQ_1—*Are CAIQs effective to compare and rank the security of CSPs?*
- RQ_2— *Are CAIQs perceived as ease to use (PEOU) to compare and rank the security of CSPs?*
- RQ_3— *Are CAIQs perceived as useful (PU) to compare and rank the security of CSPs?*

4.2. Measurements

To measure the *effectiveness* of using CAIQ, we assessed the correctness of the selection made by the participants. We asked two security experts (among the authors of this paper) to perform the same task of the participants. Then, we used the results produced by the experts as baseline to evaluate the correctness of the provider selected by the participants.

Instead, to measure the participants' *perception* of using CAIQs to select CSPs, we administered them a post-task questionnaire inspired to the Technology Acceptance Model (TAM) [34]. The questionnaire consisted of seven questions: five closed questions and two open questions: Q_1: The questions and answer in the CAIQ are clear and ease to understand (PEOU); Q_2: CAIQs make easier to assess and compare the security posture of two cloud providers (PEOU); Q_3: The use of CAIQs would reduce the effort required to compare the security posture of two cloud providers (PEOU); Q_4: The use fo CAIQs to assess and compare the security posture of two cloud provider was useful (PU); and Q_5: CAIQs do not provide an effective and complete solution to the problem of assessing and comparing the security posture of two cloud providers (PU). The closed questions were with answers on a 5 Likert scale: Strongly Agree (1) to Strongly Disagree (5).

The two open questions were included to collect insights into the rationale for selecting a CSP over another: (a) which of the two cloud providers better addresses BankGemini data protection and compliance requirements and (b) why the second provider worse addresses BankGemini security and compliance concerns.

4.3. Procedure

In order to measure the *actual effectiveness* and *perception* of using CAIQs to compare and select a cloud provider, the participants of our study were asked to impersonate BankGemini, a fictitious bank who would like to move their online banking services to the the cloud. BankGemini has very stringent requirements on data protection and legal compliance and has to select a cloud provider that meets its requirements. Due to the limited time available to run the study, we had to simplify the task for the participants. First, the participants only had to select the more secure cloud provider among only two cloud providers rather than several ones like it happens in practice. The participants were requested to choose among to real cloud providers Acquia and Capriza the one which better fulfills its data protection and compliance requirements. Second, the participants did not specify the security requirements against which comparing the two cloud providers but the requirements were given to them as part of the scenario introducing BankGemini.

4.4. Study Execution

The study consisted of three controlled experiments that took place at different locations. The first experiment took place at the University of Trento. The second one was organized at the Goethe University Frankfurt. The last experiment was conducted at University of Southampton. The same settings were applied for the execution of the three experiments. First, the participants attended one

hour lecture on cloud computing, the security and privacy issues related to cloud computing and the problem of selecting a cloud provider that meets the security needs of a tenant.

Then, 10 min were spent to introduce the participants to the high level goal of the study. The participants were explained that they had to play the role of the tenant—BankGemini—which has specific data protection and compliance requirements and that they had to select a CSP between Acquia and Capriza that better fulfils these requirements. To perform the selection, the participants were provided with:

- a brief description of BankGemini including the security requirements (for an example, refer to Appendix A)
- the CAIQ for Acquia and Capriza (see Supplementary Materials).

They were given 40 min to read the material and select the best CSP given the security requirements. After the task, they had 15 min to complete the post-task questionnaire.

4.5. Participants' Demographics

In our study we involved a total of 44 students with a different background. The first experiment conducted at the University of Trento involved 26 MSc students in Computer Science. The second one organized at the Goethe University Frankfurt involved 4 students in Business and IT. The last experiment conducted at University of Southampton had 14 MSc students in Cyber Security as participants. Table 4 highlights the background of the participants. Most of the participants (70%) had at least 2 years of working experience. Most of the participants have some knowledge in security and privacy but were not familiar with the online banking scenario that they analyzed.

Table 4. Overall Participants' Demographic Statistics.

Variable	Scale	Mean/ Median	Distribution
Education Length	Years	4.7	56.8% had less than 4 years; 36.4% had 4–7 years; 6.8% had more than 7 years
Work Experience	Years	2.1	29.5% had no experience; 47.7% had 1–3 years; 18.2% had 4–7 years; 4.5% had more than 7 years
Level of Expertise in Security	0 [1]–4 [2]	1 [3]	20.5% novices; 40.9% beginners; 22.7% competent users; 13.6% proficient users; 2.3% experts
Level of Expertise in Privacy	0 [1]–4 [2]	1 [3]	22.7% novices; 38.6% beginners; 31.8% competent users; 6.8 % proficient users
Level of Expertise in Online Banking	0 [1]–4 [2]	1 [3]	47.7% novices; 34.1% beginners; 15.9% competent users; 2.3% proficient users

[1] Novice. [2] Expert. [3] Median.

4.6. Results

In this section we report the results on the actual and perceived effectiveness of using CAIQs to compare and rank CSPs.

4.6.1. Actual Effectiveness

To evaluate the correctness of the selection made by the participants we have asked two security experts to perform the same task of the participants. The experts agreed that the provider that best meets BankGemini's security requirements is Aquia. Indeed, Aquia allows tenants to decide the location for data storage, enforces access control for tenants, cloud provider's employees and subcontractors, monitors and logs all data accesses, classify data based on their sensitivity, and clearly defines the responsibilities of tenants, cloud providers and third parties with respect to data processing, while Capriza does not.

As shown in Figure 5, the results are not consistent across the three experiments. In the first experiment, the number of participants who selected Aquia is basically the same of the one who selected Capriza. However, in the second and the third experiment almost all the participants correctly identified Aquia as the cloud provider that best satisfies the given security requirements. If look the overall results, most of the participants (68%) were able to identify the correct cloud service provider based on the CAIQ, which indicates that CAIQ could be an effective tool to comparing and ranking the security posture of CSPs.

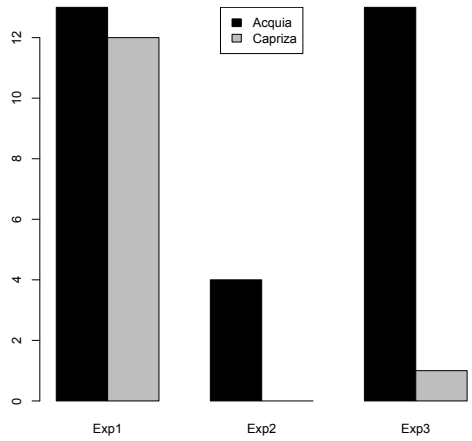


Figure 5. Actual Effectiveness—Cloud Provider Selected in the Experiments).

4.6.2. Perceived Effectiveness

Table 5 reports the mean for the answers related to PEOU and PU. The mean of the answers for all the three experiments is close to 3, which means that the participants are not confident that CAIQs make easier to compare and rank the security of CSPs and that are useful to perform the comparison and ranking of cloud service providers. These results are consistent among the three experiments. To test whether there is a statistically significant difference among the answers given by the participants in the three experiments, we run the Kruskal-Wallis statical test, the non-parametric alternative to one-way ANOVA for each question on PEOU and PU and on overall PEOU and PU. We assumed a significance level $\alpha = 0.05$. The p-values returned by Kruskal-Wallis test are reported in Table 5. The p-values are all greater than α, and therefore we have to accept the null hypotheses that there is no difference in the mean of the answers given by the participants in the three experiments. This means that all the participants believe that CAIQs are not ease to use and not useful to compare and select a cloud service provider.

Table 5. Questionnaire Analysis Results—Descriptive Statistics.

		Mean				
Q	**Type**	**Exp1**	**Exp2**	**Exp3**	**All**	***p*-Value**
Q1	PEOU	3	3.7	2.9	3.0	0.3436
Q2	PEOU	2.9	2.7	2.7	2.8	0.8262
Q3	PEOU	2.4	2.2	2.4	2.4	0.9312
Q4	PU	2.4	2.5	2.2	2.3	0.9187
Q5	PU	3.1	3.2	3.0	3.1	0.8643
PEOU		2.8	2.9	2.7	2.7	0.7617
PU		2.7	3.0	2.6	2.7	0.9927

4.7. Threats to Validity

The main threats that characterize our study are related to conclusion and external validity.

Conclusion validity is concerned with issues that affect the ability to draw the correct conclusion about the relations between the treatment and the outcome of the experiment. One possible threat to conclusion validity is related to how to evaluate the effectiveness of CAIQs in comparing and ranking the security posture of CSPS. Actual effectiveness should be assessed based on the correctness of the results produced by the participants. Therefore, in our study we asked two of the authors of this paper to perform the same selection task performed by the participants and use their results as baseline to evaluate the correctness of the best CSP identified by the participants.

External validity concerns the ability to generalize experiment results beyond the experiment settings. The main threat is related to the *use of the students instead of practitioners*. However, some studies have argued that students perform as well as professionals [35,36]. Another threat to external validity is the *realism of experimental settings*. The experiments in our study were organised as a laboratory session and therefore the participants had limited time to by the participants in comparing and ranking the security posture of CSPs. For this reason we had to simplify the task by providing to the participants Bank Gemini's security requirements, rather then letting them identify the requirements. However, this is the only simplification that we introduced. For the rest, the task is the same that a tenant would perform when selecting and comparing the security of CSPs.

4.8. Implications for Practice

The CAIQ provides a standard framework that should help tenants to assess the security posture of a CSP. The last version of the CAIQ includes 295 security controls grouped in 16 domains. Each of this control has one or more "yes, no or not applicable" control assertion questions which should allow a tenant to determine whether a provider implements security controls that suit the tenant's security requirements.

The results of our study show that the selection of a cloud provider based on the CAIQ's questions and answers could be effective because most of the participants were able to correctly select Aquia as the CSP that best meet the requirements of the tenant. However, the participants of our study are not confident that the approach is ease to use and useful to select and compare the security posture of CSPs.

The main reason why CAIQ is not perceive as ease to use and useful, is that for each CSP to be compared, a tenant has to go through 295 questions in the CAIQ, identify those questions that match the tenant security requirements, and evaluate the answers provided by the CSP to decide if the corresponding security control is supported or not. This is quite a cumbersome task for the tenant.

Therefore, there is the need for an approach that extracts from the CAIQs the information to determine if a CSP meets a tenant's security requirements and based on this information assesses the overall security posture of the provider.

5. Ranking Cloud Providers' Security

In this section we present an approach that facilitates the comparison of the security posture of CSPs based on CAIQ's answers. The approach is illustrated in Figure 6. There are three main actors involved: the tenant, the alternative CSPs, and the cloud broker. A cloud broker is an intermediary between the CSPs and the tenant, that helps the tenant to choose a provider tailored to his security needs (cf. NIST Cloud Computing Security Reference Architecture [37]). (For example Deutsche Telekom is offering this service [38]). In the setup, the broker has to assess the answers of the CSPs (classification and scoring) and define the security categories which are mapped to the CAIQ's questions. The list of security categories is then provided to the tenant. For the ranking, the broker first selects the candidate CSPs among the ones that deliver the services requested by the tenant. Then, it ranks the candidate providers based on the weighted security categories specified by the tenant and the answers that the providers gave to the CAIQ. The list of ranked CSPs is returned to the tenant, who uses the list as part of his selection process.

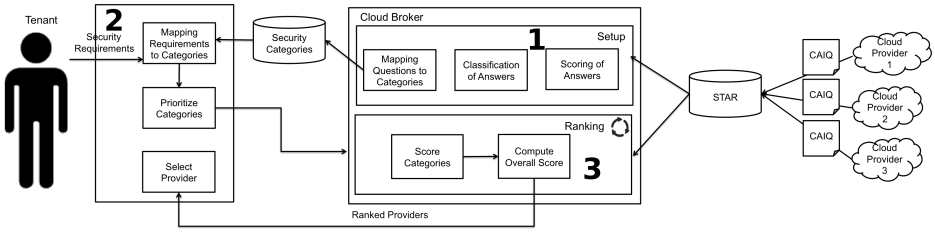


Figure 6. Security-Aware Cloud Provider Selection Approach.

The approach to rank CSPs adopts the Analytic Hierarchy Process (AHP) [26]. The first step is to decompose the selection process into a hierarchy. The top layer reflects the goal of selecting a secure CSP. The second layer denotes the security categories with respect to which the CSPs are compared while the third layer consists of the CAIQ's questions corresponding to the security categories. The bottom most layer contains the answers to the CAIQ's questions given by the different CSPs. The hierarchy is shown in Figure 7: weights and calculator symbols near each layer denote that a weight and a score for that layer is computed while the number on the symbols refer to the section in the paper were the computation is described. Similarly, the pad symbol denotes that the scores are aggregated.

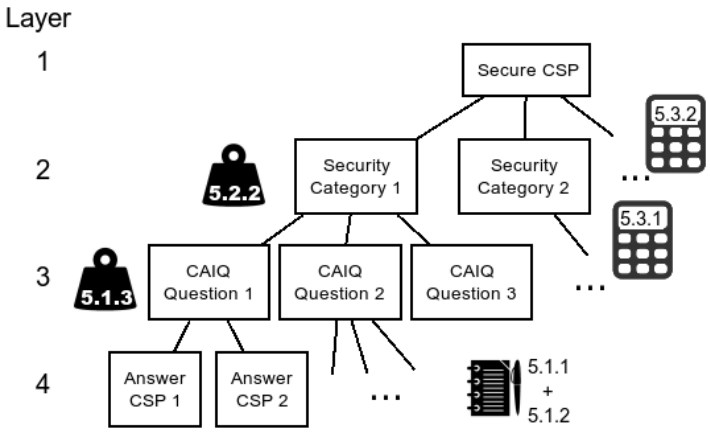


Figure 7. Hierarchies of Analytic Hierarchy Process (AHP) based Approach.

The result at the end of the decision making process is a hierarchy where each CSP gets a overall score and a score for each category. This allows the tenant not only to use the overall result in CSP selection processes with other criteria, but also to reproduce the CSPs' strengths and weaknesses regarding each category. For this reason, we chose to base our approach on AHP because it not only comes up with a result, but also provides some information on how the score was calculated (the scores of each category). This allows further reasoning or an adaptation of the requirements/scoring should the tenant not be confident with the result. In what follows we present in details each step of the CSP selection process.

5.1. Setup

Before the cloud broker can identify the optimal CSP based on the tenant's security needs there are three main steps he has to perform: classification of answers, scoring of answers and mapping questions from the CAIQ to security categories. Note that these steps have to be done only once for each provider present in the STAR.

5.1.1. Classification of Answers

The original AHP approach would require a pairwise comparison of all answers to each question. However, given the 37 (65) providers and 197 questions this would require 131202 (409760) comparisons and therefore is not feasible. Thus, the answers have to be manually classified which is extremely time consuming. The classification is reported in Table 6. Other classifications are also possible, depending on the new classification it may be sufficient to only re-rate a part of the answers.

Table 6. Possible Classes for Answers in CAIQ.

Answer	Comment Class	Description
Yes	Conflicting	The comment conflicts the answer.
Yes	Depending	The control depends on someone else.
Yes	Explanation	Further explanation on the answer is given.
Yes	Irrelevant	Comment is irrelevant to the answer.
Yes	Limitation	The answer 'yes' is limited or related due to the comment.
Yes	No comment	No comment was given.
No	Conflicting	The comment conflicts the answer.
No	Depending	The control depends on someone else.
No	Explanation	Further explanation on the answer is given.
No	Irrelevant	Comment is irrelevant to the answer.
No	No comment	No comment was given.
NA	Explanation	Further explanation on the answer is given.
NA	Irrelevant	Comment is irrelevant to the answer.
NA	No Comment	No comment was given.
Empty	No comment	No answer at all
Unclear	Irrelevant	Only comment was given and thereupon it was not possible to classify the answer as one of Y/N/NA.

The comments are used to further rate the answers of CSPs in more detailed classes. "Yes", "No" and "Not applicable" answers are assigned to the class "*No comment*" if the CSP did not give a comment. If the given answer is further described, for example, if additional information of the control in place, why the control is not in place or why this question is not applicable is given, the answers are assigned to the class "*Explanation*". If there is a comment, but it does not explain the answer of the provider, the answer is classified as "*Irrelevant*". An example for this class is the repeating of the question as a full sentence. Also comments about Non disclosure agreements which may have to be signed before were put in this class. For "yes" and "no" answers, two additional classes are considered: "*Depending*" if the provider claims that the control depends on a third party, and "*Conflicting*" if the

answer conflicts with the statement of the comment. For example "Yes, not yet started" means that either the control is not in place or the comment is wrong. For "yes" answers also the class "*Limitation*" is used when the comment limits the statement that the control is in place. Examples for this are comments which restrict the control to specified systems, which means that the control is not in place for all systems or when it is asked if the provider makes documentation available to the tenant and the comment restricts that to summaries of the specified documents. For empty answers only the class "*No comment*" is considered and for unclear answers only the class "*Irrelevant*" is used.

5.1.2. Scoring of Answers

Once the answers are classified, for each of the answers a score as to be computed to determine how the CSPs performs for each question (3rd AHP layer, sub criteria). The scoring depends on the aim the tenant wants to achieve, thus other scores are possible. For our approach we distinguish between two kind of tenants: tenants who really want to invest in security and tenants who are primarily interested in compliance (cf. Reference [39]). The tenant who wants to invest in security tries to reduce the risk of data loss. Therefore, he wants to compare the CSPs based on the risk level that incidents (e.g., loss of data, security breaches) happen. Thus, the best answer is a "Yes" with an "Explanation", followed by "Yes" answers with "No comment" or when the provider claims that the control is handled by a third party. "Irrelevant" comments, "Limitation", or even "Conflicting" comments may indicate that the control is not properly in place or not in place at all. If the provider claims that the control is not in place, the best the tenant can expect is an explanation why it is not in place, while conflicting answers may offer a chance that this control is in spite of the provider's answer in place. If the provider answered "Non Applicable", the tenant may have chosen a provider offering an unsuitable service or the provider may not have recognised that this control is relevant for him. Thus, "Non Applicable" answers were rated slightly lower than "No" answers. "Empty" and "Unclear" comments score lowest.

Instead, the tenants who are interested in compliance try to reduce the risk that if an incident occurs, there is no claim for damages or lost lawsuit. Thus, the tenant's interest is to compare the CSPs based on the risk level that he is sued after an incident has happened. Thus, basically most of the"yes" answers allow the tenant to blame his provider, should an incident have happened. However, "Limitation" and 'Conflicting" comments are scored lower, since a judge might conclude that the tenant should have noticed that. "No" answers score 0 as the latter would imply being surely not compliant. "Not applicable,""Empty" or "Unclear" answers leave at least a basis for discussions, and thus have a low score.

The scoring schemes for these two types of tenants discussed above were independently approved by three experts and are shown in Table 7.

Compared to the classification of the answers, the mapping of answer classes to scorings is less effort, but still a very decisive step which should be done by experts from the cloud broker based on the tenants' desired aims.

5.1.3. Mapping of Questions to Security Categories

The questions from CAIQ need to be mapped to security categories and assigned scores reflecting their importance to the corresponding category. This is basically the decision which sub criteria (3rd AHP layer) belong to which criteria (2nd AHP layer). Examples for security categories are: access control, data protection at rest/transport, patching policy, and penetration testing. The weight can be either given by comparing the security categories pairwise or as an absolute score.

The used score is shown in Table 8. Its range is from one to nine. If an absolute score is given (also in the range from one to nine), the relative weight for two categories (questions) may be derived by subtracting the lower score from the higher score and adding one. We give an example in the next section.

Table 7. Possible Scoring for Tenants Interested in Security or Compliance.

Answer	Comment Class	Security	Compliance
Yes	Explanation	9	9
Yes	No comment	8	9
Yes	Depending	8	9
Yes	Irrelevant	7	9
Yes	Limitation	6	7
Yes	Conflicting	5	5
No	Explanation	4	1
No	Conflicting	4	1
No	No comment	3	1
No	Depending	3	1
No	Irrelevant	2	1
NA	Explanation	3	3
NA	No comment	2	3
NA	Irrelevant	2	3
Empty	No comment	1	2
Unclear	Irrelevant	1	2

Table 8. Weights for Comparing Importance of Categories and Questions.

Weight	Explanation
1	Two categories (questions) describe an equal importance to the overall security (respective category)
3	One category (question) is moderately favoured over the other
5	One category (question) is strongly favoured over the other
7	One category (question) is very strongly favoured over the other
9	One category (question) is favoured over the other in the highest possible order

The result from this step is a list of predefined security categories and a list of weighted questions from the CAIQ mapped to the categories. The security domains provided by the CAIQ would be quite natural to use, but its use has some drawbacks. We give an additional mapping, since not every question should have the same weight inside each category. Additionally, some questions may contribute to different security categories whereas each question is part of exactly one domain in CAIQ. Furthermore, answers are not distributed equally among the different domains. Some domains essentially contain almost only questions with yes answers (cf. Figure 4). Thus, our approach is more fine-grained. We also allow different granularity, for example, for one tenant confidentiality may be sufficient, since it is only one of the tenant's multiple security requirements. Another tenant may be especially interested in that category and regard data protection at rest and data protection at transport as different security categories instead. A sample table is given in the next section (cf. Table A1).

5.2. Tenant's Task

The following steps have to be performed by the tenant, but the tenant could also be supported by experts from the cloud broker.

1. *Security Requirements:* The tenant specifies the security requirements on the data and/or applications he would like to outsource to a CSP.
2. *Map requirements to security categories:* The tenant has to map the security requirements to the predefined security categories provided by the cloud broker and assign a weight to each category that quantifies its overall importance to the tenant. The weight can be either given by comparing categories pairwise or as an absolute score. The result is a subset of the security

categories predefined by the cloud broker along with their score. This defines the 2nd layer of the AHP hierarchy.

3. *Confirming setup:* If the tenant does not agree with the choices made during the setup phase, he has to ask his cloud broker to specify an alternative version. Especially, the tenant may ask for additional predefined security categories if they do not fit his needs.

5.3. Ranking Providers

The evaluation of the previously gathered weights and scores is done bottom up by the cloud broker.

5.3.1. Scoring Security Categories

We assume, there are I security categories c_i with J_i questions each and $1 \leq i \leq I$. For each security category c_i the scores of the CSP's answers to the relevant questions q_{ij} have to be compared (with $1 \leq j \leq J_i$). We already described in Section 5.1.2 how we classified those answers. We compare them by building the difference of their scores and adding one. The interpretation of those comparison scores is shown in Table 9.

Table 9. Scores for Comparing Quality of Answers to CAIQ.

Score	Explanation
1	Two answers describe an equal implementation of the security control
3	One answer is moderately favoured over the other
5	One answer is strongly favoured over the other
7	One answer is very strongly favoured over the other
9	One answer is favoured over the other in the highest possible order

The scores are transferred to the matrix A_{ij} the following way: If their score is the same, the entry is 1 for both comparisons. For superior answers, the difference of the two scores plus one is used, for inferior answers its reciprocal is used (cf. Table 10 and Equation (1) for an example). Next, for each matrix A_{ij}, the matrix's principal right eigenvector α_{ij} is computed. For each question q_{ij} in category c_i the square matrix C_i is built from comparing the weights of the questions' importance to the corresponding category in the same way and its eigenvector γ_i is computed.

Table 10. Comparison Table.

Superior	Inferior	Comp.
CSP 1	CSP 2	x
CSP 3	CSP 1	y
CSP 3	CSP 2	z
⋮	⋮	⋮

$$\begin{pmatrix} 1 & x & \frac{1}{z} & \cdots \\ \frac{1}{x} & 1 & \frac{1}{y} & \cdots \\ z & y & 1 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} \tag{1}$$

The eigenvectors of the answers' scores α_{ij} are then combined to a matrix A_i. By multiplying A_i with the eigenvector γ_i of the questions' importance, the vector p_i is determined.

$$A_i \cdot \gamma_i = \begin{pmatrix} \alpha_1 & \alpha_2 & \dots & \alpha_{J_i} \end{pmatrix} \cdot \gamma_i = p_i, \tag{2}$$

p_i indicates each CSP's priority concerning category c_i.

5.3.2. Computing the Overall Score

The comparisons of the categories' weights as described in Section 5.2 are used to compute a matrix W analogous to the matrices representing the comparisons of the answers' quality and the questions' importance to a category. We denote its eigenvector with ω. The priorities of the categories p_i are then combined to a matrix P. By multiplying them, the overall priority p is obtained.

$$P \cdot \omega = \begin{pmatrix} p_1 & p_2 & \dots & p_{J_i} \end{pmatrix} \cdot \omega = p, \tag{3}$$

p adds up to 1 and shows the priority of CSPs' answers fulfilling the tenant's requirements.

5.4. Implementation

We have implemented our approach in the R programming language. The classifications and score of the answers and the security categories were stored in a SQL database. In the same database we also imported the CAIQ's answers from the providers. As we already discussed in Section 3.2 this is not a trivial task. From the submitted document formats, it is by far the easiest to export the data from spreadsheets (XLS) compared to text editor files (DOC) or the Portable Document Format (PDF). Referring to the different styles of answering it was easier to extract information from CAIQ version 1.1 if it had two columns or from version 3.0.1 since here answers and comments are separated. In addition, many CSPs changed the number of columns by inserting or deleting columns, and thus we needed to manually select the columns containing the CSPs' answers. Additionally some of the CSPs answered questions in blocks. This resulted either in a listing of answers in the same cell (separated with spaces or line breaks), or by answers prefixed with the control id (CID). Thus, most of the questionnaires' data could only be processed semi-automatically and had to be manually verified.

As described in Section 3.2, some of the CSPs did not provide a clear "yes/no"-answer and only had a verbal answer. To limit the impact of our interpretation of the CSPs' answers, we only processed the questionnaires where there were "yes/no"-answers to all questions or at least to most of them. For the few remaining questions without explicit answer, we derived the answer manually by examining the comment. If no comment was given, we classified the answer as "empty", if it was not possible to conclude whether the comment means, yes, no or not applicable, we classified it as "unclear". Given these restrictions, we ended up with answers from 37 CSPs for version 1.1 and 189 for version 3.0.1 in July 2017.

5.5. Implications for Practice

In this section, we introduced a novel approach to select a secure CSP, showed that it is feasible by a proof of concept implementation. Within the necessary steps some effort is needed for the setup, in particular for classifying and scoring the CSPs' answers to the CAIQ. Since this effort is only needed once, we propose that a cloud broker can offer this as a service. Besides assessing the security requirements, the most difficult task for the tenant is to map the security requirements to the security categories provided by the cloud broker and to prioritize the requirements' categories. Again, the cloud broker may offer to support the tenant and offer a (paid) service. With the requirements from the tenant and the assessment of the questionnaires, the ranking of the CSPs can be done automatically. As a last step, the tenants may select a CSP, should carefully double-check if the CSP's service level agreements are in line with the questionnaire and in particular include the requirements important to them.

If tenants are on their own terms, they suffer from the amount of different CSPs to consider and from the effort needed to classify all questionnaires. In particular, since we learned during our implementation that the assessment of the questionnaires can only be done semi-automatic, for example, for answers without a comment and many of the questionnaires and their answers have

to be processes manually. On the other hand, once the assessment is done, it can be used for multiple selection processes, so a (trusted) third party is necessary. The third party could only be avoided with additional effort either from the tenant's side or from the CSPs' side when they would be required to provide their answers in a specific machine-readable form.

6. Evaluation

In this section we assess different aspects of our approach to cloud provider ranking based on CAIQs. First of all we evaluate how ease is for the tenant to map the security categories to the security requirements and assign a score to the categories. Then, we evaluate the effectiveness of the approach with the respect to correctness of CSP selection. Last, we evaluate the performance of the approach.

Scoring of Security Categories. We wanted to evaluate how ease is for a tenant to perform the only manual step required by our approach to CSP ranking: map their security requirements to security categories and assign a score to the categories. Therefore, we asked to the same participants of the study presented in Section 4 to perform the following task. The participants were requested to map the security requirements of Bank Gemini with a provided list of security categories. For each category they were provided with a definition. Then, the participants had to assign an absolute score from 1 (not important) to 9 (very important) denoting the importance of the security category for Bank Gemini. They had 30 min to complete task and then 5 min to fill in a post task-questionnaire on the perceived ease of use of performing the task. The results of analysis of the post-task questionnaire are summarized in Table 11. Participants believe that the definition of security categories was clear and ease to understand since the mean of the answers is around 2 which corresponds to the answer "Agree". We tested the statistical significance of this result using the one sample Wilcox signed rank test setting the null hypothesis $\mu = 3$, and the significance level $\alpha = 0.05$. The p-value is <0.05 which means that result is statistically significant. Similarly, the participant agree that it was ease to assign a weight to security categories with statistical significance (one sample t-test returned p-value = 0.04069). However, they are not certain (mean of answers is 3) that assigning weights to security categories was ease for the specific case of Bank Gemini scenario. This result, though, is not statistically significant (one sample t-test returned p-value = 0.6733). Therefore, we can conclude the scoring of security categories that a tenant has to perform in our approach does not require too much effort to performed.

Table 11. Scoring of Categories Questionnaire—Descriptive Statistics.

Type	ID	Questions	Mean	Median	sd	p-Value
PEOU	Q_1	In general, I found the definition of security categories clear and ease to understand	2.29	2	0.93	6.125×10^{-5}
PEOU	Q_2	I found the assignment of weights to security categories complex and difficult to follow	3.4	4	1.4	0.04069
PEOU	Q_3	For the specific case of the Home Banking Cloud-Based Service it was ease to assign weights to security categories	3.06	3	1.06	0.6733
Overall PEOU			2.91	3	1.13	0.3698

Effectiveness of the Approach. To evaluate the correctness of our approach, we determined if the overall score assigned by our approach to each CSP reflects the level of security provided by the CSPs and thus if our approach leads to select the most secure CSP. For this reason we used the three scenarios from our experiment and additionally created a more complicated test case based on the FIPS200 standard [40]. The more sophisticated example makes use of the full CAIQ version 1.1 (197 questions) and comes up with 75 security categories. As we did for the results produced by the participants of our experiments, we have compared the results produced by our approach for the three scenarios and the additional test case with the results produced by the three experts on the same scenarios. Our approach results were consistent with the results of the experts. Furthermore, the results of the 17 participants

who compared two CSPs by answers and comments on 20 questions, are also in accordance to the result of our approach.

Performance. We evaluated the performance of our approach with respect to the number of providers to be compared and the number of questions used from the CAIQ. For that purpose we generated two test cases. The first test case is based on the banking scenario that we used to run the experiment with the students. It consists of 3 security requirements, 20 CAIQ's questions and 5 security categories. The second test case is the one based on the FIPS200 standard and described above (15 security requirements, 197 questions, 75 security categories). We first compared only 2 providers as in the experiment and then compared all the 37 providers in our data set for version 1.1. The tests were run on a laptop with an Intel(R) Core(TM) i7-4550U CPU. Table 12 reports the execution time of our approach. It shows the execution time for ranking the providers (cf. Section 5.3) and the overall execution time, which also includes the time to load some libraries and query the database to fetch the setup information (cf. Sections 5.1 and 5.2).

Table 12. Performance Time of Our Approach as a Function of the Number of CSP and the Number of Questions.

N° CSP	N° Questions	N° Categories	Ranking	Total
2	20	5	∼0.5 s	<1 s
37	20	5	48 s	50 s
2	197	75	1 min 50 s	<2 min
37	197	75	34 min	<35 min

Our approach takes 35 min to compare and rank all 37 providers from our data based on a full CAIQ version 1.1. This is quite fast compared to our estimation that the participants of our experiment would need 80 min to manually compare only two providers with an even easier scenario. This means that our approach makes it feasible to compare CSPs based on CAIQ's answers. Another result is that as expected the execution time increases with the number of CSPs to be compared, the number of questions and the number of security categories. This execution time could be further reduced if the ranking of each security category would be run in parallel rather then sequentially.

Feasibility. The setup of this approach requires some effort, which need only to be rendered once. Therefore, it is not feasible for the tenants to do the set-up for a single comparison and ranking. However, if the comparison and ranking is offered as a service by a cloud broker, and thus is used for multiple queries, the set-up share of the effort decreases. Alternatively, a third party such as the Cloud Security Alliance could provide the needed database to the tenants and enable them do to their own comparisons.

Limitations. Since security cannot be measured directly, our approach is based on the assumption that the implementation of the controls defined by the CCM is related to security. Should the CCM's controls fail to cover some aspects or be not related to the security of the CSPs the result of our approach would be effected. Additionally, our approach relies on the assumption that the statements given in the CSPs' self-assessments are correct. The results would be more valuable, if all answers would have been audited by an independent trusted party and certificates were given, but unfortunately as of today this is only the case for a very limited number of CSPs.

Evolving CAIQ versions. While our approach is based on CAIQ version 1.1, it is straight forward to run it on version 3.0.1 respectively version 3.1 also. However, with different versions in use cross version comparisons can only be done with the overlapping common questions. We provide a mapping between the 169 overlapping questions for version 1.1 and 3.0.1 (cf. Section 3.1). If CAIQ version 1.1 will no longer be used or the corresponding providers are not of interest, the mappings of the questions to the security categories may be enhanced to make use of all 295 questions of CAIQ version 3.0.1.

7. Conclusions and Future Work

In this paper we investigated the issues related to CSP selection based on the CSPs' self-assessments and their answers to the Consensus Assessments Initiative Questionnaire (CAIQ). We have discussed first the issues related to processing the CAIQ, namely many CSPs did not follow a standard format to answer the questionnaire and some CSPs did not provide clear answers on which controls they support. Therefore, to facilitate the automatic data processing of CAIQ it would be helpful to have a more standardized data set with unambiguous statements. This could either be a simple text-based format like Comma Separated Variable files (CSV) or an XML-based format like a to be defined Cloud Service Security Description Language or a Multi-Criteria Decision Analysis Modelling Language such as XMCDA [41].

Given these issues we have conducted a controlled experiment with master students to assess whether manually selecting the CSP that best meets the security requirements of a tenant based on the answers to CAIQ is feasible in practice. The experiment revealed that such an approach is not feasible in practice. In fact, the participants took approximately eight minutes to compare two providers based on the answers given to a small subset (20 questions) of the questions included in the CAIQ. If we scale to the full questionnaire which contains around 200 questions, a tenant would take around one and a half hours to compare just two cloud providers.

For this reason, we have proposed an approach that facilitates a tenant in the selection of a provider that best meets its security requirements. The tenant has only to identify the security requirements, rank them, and assign them to predefined security categories. Then the cloud broker uses the Analytic Hierarchy Process to compute a score for each security category based on the answers given by the providers to corresponding questions in the CAIQ. The output is a ranked list based on the weighted overall score for each provider as well as each provider's ranking for each security category. Our approach is quite flexible and allows to be easily customized should the tenant want to change the included scoring, categories or mappings to his own needs.

An preliminary evaluation of the actual efficiency of the approach shows that it takes roughly a minute per provider to compare and rank CSPs based on the full CAIQ.

We are planning to extend our work in four main directions:

Classification of Answers and Questions. The classification of answers and questions are key steps in our approach for selecting CSPs but are also very time consuming. To automatize these steps we will use machine learning techniques to build a text classifier that automatically associates answers and questions to the corresponding class.

Visualization. We focused on providing input for a general CSP selection approach. However, it may be helpful to display the results of the selection process to the tenant. A simple idea could be to build an interface that follows the traffic light metaphor: for each category in the CAIQ it shows in green the categories that satisfy the security requirements of the tenant, in red the one that are not fulfilled and in grey the one that are not relevant with respect to the tenant's security requirements.

Measuring Security. Since security can not be measured directly we focused on experts' judgement to evaluate our approach. It would be interesting to conduct a standardized penetration testing for a couple of the CSPs and match the results with the providers' answers to the CAIQ.

Supplementary Materials: The following are available online at http://www.mdpi.com/2078-2489/11/5/261/s1.

Author Contributions: Conceptualization, F.M., S.P., F.P., and J.J.; methodology, F.M., S.P., F.P.; software, S.P.; validation, S.P., F.P., J.J., and F.M.; investigation, S.P. and F.P.; resources, S.P., F.P., and F.M.; data curation, S.P. and F.P.; writing–original draft preparation, S.P. and F.P.; writing–review and editing, F.M. and J.J.; visualization, S.P.; supervision, F.M. and J.J.; funding acquisition, J.J. and F.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partly funded by the European Union within the projects Seconomics (grant number 285223), ClouDAT (grant number 300267102) and CyberSec4Europe (grant number 830929).

Acknowledgments: We thank Woohyun Shim for fruitful discussions on the economic background of this paper and Katsiaryna Labunets for her help in conducting the experiment.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Example: Application of Our Approach

To illustrate our approach, we show how it is applied to the banking scenario we used in the controlled experiment described in Section 4.

Appendix A.1. Setup

The classification and scoring of answers as described in the previous section meets the fictitious tenant's needs. Since the tenant is interested in security, the corresponding scoring for security mentioned in Section 5.1.1 was chosen.

The mapping of questions to security categories along with their importance to the respective category is shown in Table A1.

Table A1. Weighted Mapping from Questions to Categories.

Number	CID	Weight	Category
1	IS-03.1	7	Privacy
1	IS-03.1	7	Confidentiality
2	IS-03.2	7	Confidentiality
2	IS-03.2	7	Privacy
3	IS-03.3	3	Confidentiality
3	IS-03.3	3	Privacy
4	IS-08.1	9	Confidentiality
5	IS-08.2	9	Confidentiality
6	IS-18.1	9	Key Management
7	IS-18.2	9	Key Management
8	IS-19.1	9	Confidentiality
9	IS-19.2	9	Confidentiality
10	IS-19.3	5	Key Management
11	IS-19.4	7	Key Management
12	IS-22.1	7	Availability
⋮	⋮	⋮	⋮
20	SA-14.1	5	Integrity

Appendix A.2. Tenant's Task

The following security requirements were assumed from the description of the scenario:

- The cloud provider should protect the confidentiality of data during transport and at rest
- The cloud provider should protect the privacy of the accounting data
- The cloud provider should protect the integrity of data during transport and at rest
- The cloud provider should guarantee the availability of accounting applications and data

Based on the requirements the following predefined security categories (weights in brackets) were chosen: Confidentiality (9), Privacy (9), Integrity (9), Availability (9), and Key Management (5).

Appendix A.3. Ranking Providers

Appendix A.3.1. Scoring Security Categories

We report here only the computation of the score for the security category "Key Management" (i = 5). The score for the other categories can be computed in a similar way. For "Key Management" questions 6, 7, 10, and 11 are relevant. The scoring of the providers' answers is shown in Table A2.

Table A2. Scorings of CSPs for Questions Relevant for Key Management.

Number	Weight	CSP A	CSP B
6	9	3	4
7	9	3	4
10	5	7	7
11	7	8	7

$$A_{51} = \begin{pmatrix} 1 & 0.5 \\ 2 & 1 \end{pmatrix} \tag{A1}$$

$$A_{53} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \tag{A2}$$

For the first question (number 6, j = 1), the difference between the two scorings is one in favour to CSP B, thus the result for the comparison matrix A_{51} shown in Equation A1. The resulting matrix's principal right eigenvector is shown in Equation (A3). In the same manner, the weights of the questions are compared, a (4×4)-matrix is built and its resulting eigenvector γ_5 is left multiplied. So the priority p_5 for category c_5 ends in 0.395 versus 0.605 in favour of CSP B.

$$\begin{pmatrix} 0.391 & 0.391 & 0.0675 & 0.151 \end{pmatrix} \begin{pmatrix} 0.333 & 0.667 \\ 0.333 & 0.667 \\ 0.500 & 0.500 \\ 0.667 & 0.333 \end{pmatrix} = \begin{pmatrix} 0.395 & 0.605 \end{pmatrix} \tag{A3}$$

In the same manner, the priorities for the other security categories are determined resulting in P shown in Equation (A4).

Appendix A.3.2. Computing the Overall Score

From the weights of the categories the eigenvector ω is computed in the same manner. The result of the multiplication $P \cdot \omega$ (see Equation (A4)) delivers the overall score. The result favours CSP B with roughly 60:40 over CSP A regarding the banking scenario. In the supplementary material the result for all 37 providers for all three scenarios is given.

$$\begin{pmatrix} 0.358 & 0.606 & 0.319 & 0.292 & 0.395 \\ 0.642 & 0.394 & 0.681 & 0.708 & 0.605 \end{pmatrix} \begin{pmatrix} 0.238 \\ 0.238 \\ 0.238 \\ 0.238 \\ 0.048 \end{pmatrix} = \begin{pmatrix} 0.394 \\ 0.606 \end{pmatrix} \tag{A4}$$

References

1. NIST Special Publication 800-53—Security and Privacy Controls for Federal Information Systems and Organizations. Available online: http://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-53r4.pdf (accessed on 31 March 2020).
2. KPMG. 2014 KPMG Cloud Survey Report. Available online: http://www.kpmginfo.com/EnablingBusinessInTheCloud/downloads/7397-CloudSurvey-Rev1-5-15.pdf#page=4 (accessed on 31 March 2020).
3. Böhme, R. Security Metrics and Security Investment Models. Advances in Information and Computer Security. In Proceedings of the 5th International Workshop on Security, IWSEC 2010, Kobe, Japan, 22–24 November, 2010; Echizen, I., Kunihiro, N., Sasaki, R., Eds.; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2010, Volume 6434, pp. 10–24.

4. Akerlof, G.A. The Market for 'Lemons': Quality Uncertainty and the Market Mechanism. *Q. J. Econ.* **1970**, *84*, 488 – 500. [CrossRef]
5. Tirole, J. Cognition and Incomplete Contracts. *Am. Econ. Rev.* **2009**, *99*, 265–294. [CrossRef]
6. Pape, S.; Stankovic, J. An Insight into Decisive Factors in Cloud Provider Selection with a Focus on Security. Computer Security. In Proceedings of the ESORICS 2019 International Workshops, CyberICPS, SECPRE, SPOSE, ADIoT, Luxembourg, 26–27 September, 2019; Katsikas, S., Cuppens, F., Cuppens, N., Lambrinoudakis, C., Kalloniatis, C., Mylopoulos, J., Antón, A., Gritzalis, S., Pallas, F., Pohle, J., et al., Eds.; Revised Selected Papers; Lecture Notes in Computer Science; Springer International Publishing: Cham, Switzerland, 2019; Volume 11980, pp. 287–306.
7. Anastasi, G.; Carlini, E.; Coppola, M.; Dazzi, P. QBROKAGE: A Genetic Approach for QoS Cloud Brokering. In Proceedings of the 2014 IEEE 7th International Conference on Cloud Computing (CLOUD), Anchorage, AK, USA, 27 June–2 July 2014; pp. 304–311.
8. Ngan, L.D.; Kanagasabai, R. OWL-S Based Semantic Cloud Service Broker. In Proceedings of the 2012 IEEE 19th International Conference on Web Services (ICWS), Honolulu, HI, USA, 24–29 June 2012; pp. 560–567.
9. Sim, K.M. Agent-Based Cloud Computing. *Serv. Comput. IEEE Trans.* **2012**, *5*, 564–577.
10. Wang, P.; Du, X. An Incentive Mechanism for Game-Based QoS-Aware Service Selection. In *Service-Oriented Computing*; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2013; Volume 8274, pp. 491–498.
11. Karim, R.; Ding, C.; Miri, A. An End-to-End QoS Mapping Approach for Cloud Service Selection. In Proceedings of the 2013 IEEE Ninth World Congress on Services (SERVICES), Santa Clara, CA, USA, 28 June–3 July 2013; pp. 341–348.
12. Sundareswaran, S.; Squicciarini, A.; Lin, D. A Brokerage-Based Approach for Cloud Service Selection. In Proceedings of the 2012 IEEE 5th International Conference on Cloud Computing (CLOUD), Honolulu, HI, USA, 24–29 June 2012; pp. 558–565. [CrossRef]
13. Ghosh, N.; Ghosh, S.; Das, S. SelCSP: A Framework to Facilitate Selection of Cloud Service Providers. *IEEE Trans. Cloud Comput.* **2014**, *3*, 66–79. [CrossRef]
14. Costa, P.; Lourenço, J.; da Silva, M. Evaluating Cloud Services Using a Multiple Criteria Decision Analysis Approach. In *Service-Oriented Computing*; Springer: Berlin/Heidelberg, Germany, 2013; Volume 8274, pp. 456–464.
15. Garg, S.; Versteeg, S.; Buyya, R. SMICloud: A Framework for Comparing and Ranking Cloud Services. In Proceedings of the 2011 Fourth IEEE International Conference on Utility and Cloud Computing (UCC), Melbourne, Australia, 5–8 December 2011; pp. 210–218. [CrossRef]
16. Patiniotakis, I.; Rizou, S.; Verginadis, Y.; Mentzas, G. Managing Imprecise Criteria in Cloud Service Ranking with a Fuzzy Multi-criteria Decision Making Method. In *Service-Oriented and Cloud Computing*; Lau, K.K., Lamersdorf, W., Pimentel, E., Eds.; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2013; Volume 8135, pp. 34–48.
17. Wittern, E.; Kuhlenkamp, J.; Menzel, M. Cloud Service Selection Based on Variability Modeling. In *Service-Oriented Computing*; Liu, C., Ludwig, H., Toumani, F., Yu, Q., Eds.; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2012; Volume 7636, pp. 127–141.
18. Habib, S.M.; Ries, S.; Mühlhäuser, M.; Varikkattu, P. Towards a trust management system for cloud computing marketplaces: using CAIQ as a trust information source. *Secur. Commun. Netw.* **2014**, *7*, 2185–2200. [CrossRef]
19. Mouratidis, H.; Islam, S.; Kalloniatis, C.; Gritzalis, S. A framework to support selection of cloud providers based on security and privacy requirements. *J. Syst. Softw.* **2013**, *86*, 2276–2293. [CrossRef]
20. Akinrolabu, O.; New, S.; Martin, A. CSCCRA: A novel quantitative risk assessment model for cloud service providers. In Proceedings of the European, Mediterranean, and Middle Eastern Conference on Information Systems, Limassol, Cyprus, 4–5 October 2018; pp. 177–184.
21. Mahesh, A.; Suresh, N.; Gupta, M.; Sharman, R. Cloud risk resilience: Investigation of audit practices and technology advances-a technical report. In *Cyber Warfare and Terrorism: Concepts, Methodologies, Tools, and Applications*; IGI Global: Hershey, PA, USA, 2020; pp. 1518–1548.
22. Bleikertz, S.; Mastelic, T.; Pape, S.; Pieters, W.; Dimkov, T. Defining the Cloud Battlefield—Supporting Security Assessments by Cloud Customers. In Proceedings of the IEEE International Conference on Cloud Engineering (IC2E), San Francisco, CA, USA, 25–27 March 2013; pp. 78–87. [CrossRef]

23. Siegel, J.; Perdue, J. Cloud Services Measures for Global Use: The Service Measurement Index (SMI). In Proceedings of the 2012 Annual SRII Global Conference (SRII), San Jose, CA, USA, 24–27 July 2012; pp. 411–415. [CrossRef]
24. Cloud Services Measurement Initiative Consortium. *Service Measurement Index Version 2.1*; Technical Report; Carnegie Mellon University: Pittsburgh, PA, USA, 2014.
25. Cloud Services Measurement Initiative Consortium. Available online: https://www.iaop.org/Download/Download.aspx?ID=1779&AID=&SSID=&TKN=6a4b939cba11439e9d3a (accessed on 31 March 2020).
26. Saaty, T.L. *Theory and Applications of the Analytic Network Process: Decision Making with Benefits, Opportunities, Costs, and Risks*; RWS Publications: Pittsburgh, PA, USA, 2005.
27. Saaty, T.L. Decision making with the analytic hierarchy process. *Int. J. Serv. Sci.* **2008**, *1*, 83–98. [CrossRef]
28. Buckley, J.J. Ranking alternatives using fuzzy numbers. *Fuzzy Sets Syst.* **1985**, *15*, 21–31. [CrossRef]
29. Chang, D.Y. Applications of the extent analysis method on fuzzy AHP. *Eur. J. Oper. Res.* **1996**, *95*, 649–655. [CrossRef]
30. Cloud Security Alliance. Available online: https://cloudsecurityalliance.org/ (accessed on 31 March 2020).
31. Cloud Security Alliance. Cloud Controls Matrix. v3.0.1. Available online: https://cloudsecurityalliance.org/research/cloud-controls-matrix/ (accessed on 31 March 2020).
32. Cloud Security Alliance. Consensus Assessments Initiative Questionnaire. v3.0.1. Available online: https://cloudsecurityalliance.org/artifacts/consensus-assessments-initiative-questionnaire-v3-1/ (accessed on 31 March 2020).
33. Levenshtein, V.I. Binary codes capable of correcting deletions, insertions and reversals. *Sov. Phys. Dokl.* **1966**, *10*, 707–710.
34. Davis, F.D. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Q.* **1989**, *13*, 319–340. [CrossRef]
35. Svahnberg, M.; Aurum, A.; Wohlin, C. Using Students As Subjects—An Empirical Evaluation. In *Proceedings of the Second ACM-IEEE International Symposium on Empirical Software Engineering and Measurement*; ACM: New York, NY, USA, 2008; pp. 288–290.
36. Höst, M.; Regnell, B.; Wohlin, C. Using Students As Subjects: A Comparative Study of Students and Professionals in Lead-Time Impact Assessment. *Empir. Softw. Eng.* **2000**, *5*, 201–214. [CrossRef]
37. NIST Cloud Computing Security Working Group. *NIST Cloud Computing Security Reference Architecture*; Technical Report; National Institute of Standards and Technology: Gaithersburg, MD, USA, 2013.
38. Deutsche Telekom. Cloud Broker: Neues Portal von T-Systems lichtet den Cloud-Nebel. Available online: https://www.telekom.com/de/medien/medieninformationen/detail/cloud-broker-neues-portal-von-t-systems-lichtet-den-cloud-nebel-347356 (accessed on 31 March 2020).
39. Schneier, B. Security and compliance. *Secur. Priv. IEEE* **2004**, *2*. [CrossRef]
40. National Institute of Standards and Technology. Minimum Security Requirements for Federal Information and Information Systems (FIPS 200). Available online: http://csrc.nist.gov/publications/fips/fips200/FIPS-200-final-march.pdf (accessed on 31 March 2020).
41. Decision Deck. The XMCDA Standard. Available online: http://www.decision-deck.org/xmcda/ (accessed on 31 March 2020).

B.9 LiSRA: Lightweight Security Risk Assessment for Decision Support in Information Security

Christopher Schmitz and Sebastian Pape. LiSRA: Lightweight security risk assessment for decision support in information security. *Computers & Security*, 90, 2020. doi: 10.1016/j.cose.2019.101656. URL https://www.sciencedirect.com/science/article/pii/S0167404819301993

LiSRA: Lightweight Security Risk Assessment for Decision Support in Information Security

Christopher Schmitz and Sebastian Pape

Goethe University Frankfurt, Germany
{christopher.schmitz, sebastian.pape}@m-chair.de

Abstract

Information security risk assessment frameworks support decision-makers in assessing and understanding the risks their organisation is exposed to. However, there is a lack of lightweight approaches. Most existing frameworks require security-related information that are not available and that are very challenging to gather. So they are not suitable in practice, especially for small and medium-sized enterprises (SMEs) who often lack in data and in security knowledge. On the other hand, other explicit SME approaches have far less informative value than the proposed framework. Moreover, many approaches only provide extensive process descriptions that are challenging for SMEs. In order to overcome this challenge, we propose LiSRA, a lightweight, domain-specific framework to support information security decision-making. It is designed with a two-sided input where domain experts initially provide domain-specific information (e.g. attack scenarios for a specific domain), whereupon users can focus on specifying their security practices and organisational characteristics by entering information that many organisations have already collected. This information is then linked to attack paths and to the corresponding adverse impacts in order to finally assess the total risk. Moreover, LiSRA can be used to get transparent recommendations for future security activities and presents detailed insights on the mitigating effects of each recommendation. The security activities are being evaluated taking into account the security activities already in place, and also considering the dependencies between multiple overlapping activities that can be of complementary, substitutive or dependent nature. Both aspects are ignored by most existing evaluation approaches which can lead to an over-investment in security. A prototype has been implemented, and the applicability of the framework has been evaluated with performance and robustness analyses and with initial qualitative evaluations.

Keywords: security risk assessment, decision support, attack trees, maturity levels, security controls, ISO/IEC 27001

1. Introduction

Frameworks for information security risk assessment play a major role in the daily routines of decision-makers in information security. They are used to systematically assess the organisational security risk and to better understand the risks an organisation is exposed to. A solid risk assessment also builds the basis for an information security management system (ISMS). Otherwise, decision-makers will not be able to allocate their finite resources efficiently.

However, security risk assessment is a challenging task that normally requires a deep understanding of the relevant attack scenarios and technical knowledge about the mitigating effects of all the implemented security measures in the organisation. This poses a challenge especially for small and medium-sized enterprises (SMEs) that often do not have the capacities to run a fully-fledged information security department. Due to smaller IT budgets they often have a lack in security expertise and security-related data. Thus, most information security risk assessment frameworks are not suitable for them. Although there exist explicit SME approaches they have far less informative value than the proposed framework [1, 2]. Besides that, many approaches only present extensive process descriptions and guidelines that are challenging for SMEs [3, 4].

To address these issues, we propose LiSRA, a lightweight, domain-specific framework for decision support in information security. LiSRA is designed with a particular focus on the special needs for SMEs. Therefore, a key requirement is to mainly use already existing data and to keep the user's input to a minimum but to en-

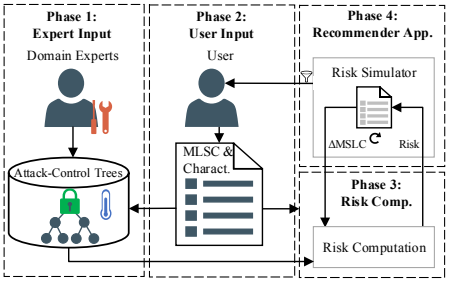


Figure 1: Overview

sure good analysis results at the same time. To meet the requirements, LiSRA expects input from both users and domain experts who are associated with the platform provider (that hosts the proposed LiSRA framework as a web-based application). The general concept is illustrated in Fig. 1. The framework assumes that organisations within a particular domain are basically exposed to similar attacks[1]. Domain experts with in-depth security knowledge for a particular domain (e. g. the electric sector) initialise the framework by providing domain-specific information (e. g. attack scenarios for a specific domain) so the user can concentrate on completing an easy-to-answer questionnaire to specify the implementation status of their security practices (that are represented by security controls).

For many organisations this only causes little extra effort because they have already collected these information. LiSRA links this information with attack trees – a well-known formalism to represent attack scenarios in a tree-based structure where high-level attack goals are decomposed into attack steps using an AND–OR tree structure [6]. They are used to calculate to which degree the implemented security controls protect against a set of attack scenarios in order to finally assess the scenario risks as well as the total risk. LiSRA can also be used to get transparent recommendations for future security activities that also provide detailed insights on their mitigating effects and how to implement them in an effective way. The term "*security activity*" is used in the sense of increasing the maturity level of a security control. Most existing approaches evaluate new security activities in isolation of security actitivies already in place, and they ignore that multiple overlapping activities can be of complementary, substitutive, or dependent nature which leads to an over-investment in security measures [7]. LiSRA explicitly addresses both aspects without bothering the user.

To further ease the data entering for the users the framework has been integrated into a web-based security management platform which eases the burden of going through a longer questionnaire. This is achieved for example by making use of small modules that are spread across the platform. They allow the users to complete or update the data needed for the risk assessment along the way when interacting with other parts of the platform [8]. Alternatively, if the data is already digitally available, e.g. as the output of an ISMS, it can also be easily imported.

The remainder of this paper is organised as follows. Section 2 presents the LiSRA framework along with a brief description of its implementation. In Section 3 an example is shown which demonstrates the framework's ease of use in the electric sector as an exemplary domain. Section 4 presents the evaluation and reports about limitations, Section 5 presents the related work, and Section 6 finally concludes and points out future research ideas.

2. LiSRA: Lightweight Security Risk Assessment

LiSRA is a lightweight security risk assessment framework for decision support in information security aiming to overcome the mentioned challenges. It models the organisation's security activities in a lightweight manner and links them with attack scenarios and their adverse impacts in order to measure the security risks. This approach can also be used to identify beneficial future security activities taking into account the effects of overlapping security activities. The framework consists of four phases:

a) *Phase 1: Expert Input.* In the first phase domain experts initially set up the framework for particular domains (e.g. the electric sector) by constructing parameterised attack trees that are linked to security controls. In a later step the user can select the domain in which his organisation operates so that the risk assessment only considers attack trees that are relevant for the respective domain. The required steps for this are illustrated in the flow chart depicted in Fig. 3 and are further described in Sect. 2.1.

b) *Phase 2: User Input.* The only user inputs required are the maturity levels of the organisation's secu-

[1]The National Electric Sector Cybersecurity Organization Resource (NESCOR) [5] for example gives an overview of domain-specific attack scenarios for the electric sector

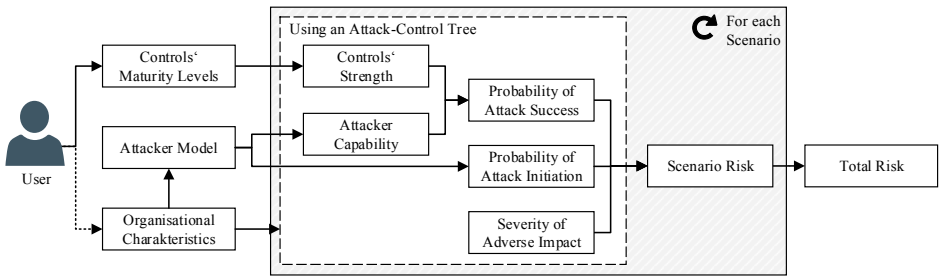


Figure 2: General Risk Computation Process

rity controls. They are used to model the implemented security practices of the organisation in a lightweight manner (see Sect. 2.2).

c) *Phase 3: Risk Computation.* Before the risk computation can start the control dependencies are resolved. This is needed because the effective maturity levels may be lower than the actual maturity levels due to control dependencies. The general risk computation process is illustrated in Fig. 2. First, the total risk is derived from scenario risks that are calculated based on both the probability of adverse impact and its severity. The probability of adverse impact is the probability that an attack is initiated and succeeds. Both factors are calculated using attack trees. The details are explained in Sect. 2.3.

d) *Phase 4: Recommender Application.* The recommender application identifies the most effective and the most cost-efficient security activities. Further information is provided in Sect. 2.4.

Since LiSRA deals with attacker behaviour, assumptions with respect to the attacker model have to be made. Here, a rational attacker is assumed to follow a best-shot strategy and always chooses the attacks and attack steps maximising his utility (according to pre-defined attacker models).

2.1. Phase 1: Expert Input

In phase 1 experts initially set up the framework for a particular domain. They gather relevant attack scenarios, transform them into a tree structure (attack trees) and link them with the respective security controls. This tree-based structure is defined as attack-control tree (ACTree) that enables determining to which extent the implemented security controls protect against attack scenarios and their associated adverse impacts. Finally, the attack-control trees are parameterised in such a way that they reflect the efficacy of controls and the attack costs. The required actions are illustrated in Fig. 3 and are described in detail in the following sections.

To make sure the system is up-to-date experts update the data at regular intervals (e.g. once per quarter) and also irregularly if the threat situation has changed significantly.

2.1.1. Identifying Attack Scenarios

The very first step is to identify the relevant attack scenarios (see A1 in Fig. 3). Experts identify both domain-independent scenarios (general attacks like malware or phishing attacks) and specific scenarios (e. g. attacking smart meters for the electric sector) for all domains that should be covered. The user can later select the domain in which his organisation operates so that the risk assessment only takes into account the relevant attack scenarios. So each domain-specific scenario has to be explicitly linked to one or more domains. It is essential to identify scenarios for both domain-specific scenarios (e. g. attacking smart meters for the electric sector) as well as for domain-independent scenarios (general attacks like malware or phishing attacks) because all organisations are exposed to general attack.

Domain experts typically already have a collection of attack scenarios because most risk assessment approaches in security management are scenario-based. So the processes will in most cases not take much time.

2.1.2. Assessing Adverse Impact

The next step is to assess the scenario's adverse impact, $I_s \in [0, 1]$ (see A2 in Fig. 3). The impact assessment is an essential factor in risk computation because it reflects the probable loss that can be expected by an attack scenario. In common and widely used frameworks

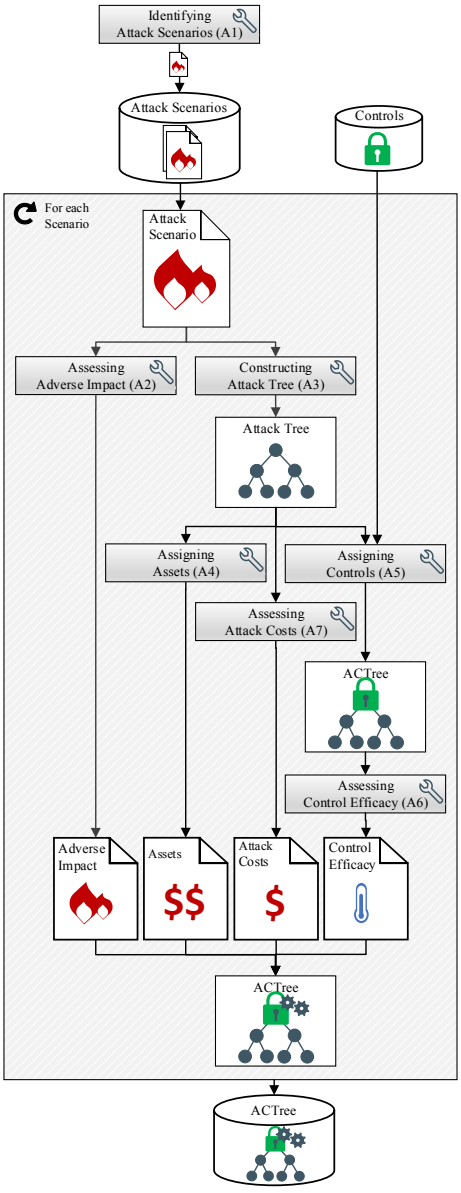


Figure 3: Expert Input – Constructing Attack-Control Trees

(such as ISO/IEC 27005 [9] and NIST SP 800-30 [10]) the impact is assessed with similar 5-point scales. Using scales the experts are familiar with eases the process of impact assessment and supports the reliability of the expert input. Besides that, it also improves the input accuracy. Therefore, the proposed LiSRA framework also uses a 5-point scale for the impact assessment. It uses the NIST impact assessment scale because it is an established scale that also contains a textual description for each impact level (in contrast to the scale used in ISO/IEC 27005). The scale is depicted in Tab. 1. For several domains there exist domain-specific, scenario-based impact assessment methods. It can make sense to combine LiSRA with one of those methods in order to further refine the analysis results.

2.1.3. Constructing Attack Trees

The identified attack scenarios are then transformed into attack trees (see A3 in Fig. 3). Attack trees are an established method in threat and risk analysis to systematicly analyse possible attack paths [11, 12]. They decomposed a high-level attack goals into single attack steps using logical AND–OR operations. Kordy et al. give a structured overview of the numerous existing variations [13].

Before constructing the trees from scratch it is recommended to follow best practices on model creation and to make use of attack pattern libraries or shared attack trees. The TREsPASS project, for instance, addressed these topics [14]. Furthermore, NESCOR provides a list of common subtrees (such as "Threat agent gains access to network") that can easily be integrated into general attack scenarios [15].

The attack trees used by LiSRA basically follow the definition of the defence trees introduced by Bistarelli et al. in 2006 [16]. The only difference is that the attack trees are extended by security controls[2] instead of concrete security measures. This modification is needed to be able to link the user's implementation status for specific security activities (defender perspective) with attacker activities (attacker perspective). For this, the vast amount of possible security measures had to be reduced by using an assessable number of roughly more than 100 security controls.

Similar to the approach by Bistarelli et al., all attacker activities are represented in leaf nodes. This does not pose a limitation because other attack tree representations where the attacker activities (and therefore the attack costs) are located in inner nodes (like ACTs by Roy

[2]A security control describes a set of security measures for the fulfillment of a security requirement.

Table 1: Impact Assessment Scale based on NIST [10]

Qualitative Values	Quantitative Values	Description
Severe	1	The attack event could be expected to have multiple severe or catastrophic adverse effects on organizational operations, organizational assets, individuals, other organizations, or the Nation.
Major	0.8	The attack event could be expected to have a severe or catastrophic adverse effect on organizational operations, organizational assets, individuals, other organizations, or the Nation.
Medium	0.5	The attack event could be expected to have a serious adverse effect on organizational operations, organizational assets, individuals other organizations, or the Nation.
Minor	0.2	The attack event could be expected to have a limited adverse effect on organizational operations, organizational assets, individuals other organizations, or the Nation.
Negligible	0	The attack event could be expected to have a negligible adverse effect on organizational operations, organizational assets, individuals other organizations, or the Nation.

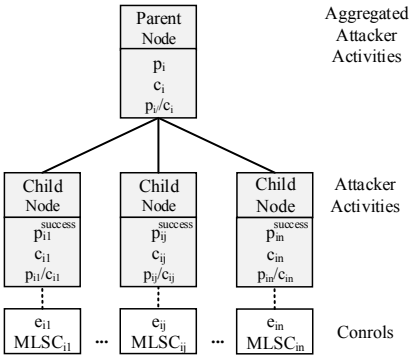


Figure 4: Parameter Notation for Attack-Control Trees

et al. [17] can easily be transformed into this representation.

The parameter notation for ACTrees, that are used in the following, is illustrated in Fig. 4 which visualises the nodes' parameters and indices. Each parent node i has a set of child nodes $j \in J$. This also holds for an attacker activity i that is assigned to j controls.

2.1.4. Assigning Assets

When the attack trees are constructed assets are assigned to corresponding nodes in the ACTree (see A4 in Fig. 3). The mapping between assets and attack steps enables to know which attacks or attack steps require the presence of which assets to be successfully performed. It is used in a later step to individualise the attack trees. We focus on the supporting assets according to ISO/IEC 27005 because "these assets have vulnerabilities that are exploitable by threats aiming to impair the primary assets of the scope (processes and information)" [9]. So attackers have to attack these "supporting assets" in the first place in order to achieve their attack goal. Therefore, an organisation that does not work with a specific asset class is not exposed to the corresponding attacks. For example, an organisation that does no work with respectively does not store any (sensitive) information on the asset class "database server" is not exposed to the attack "data theft through SQL injection". The attack step "data theft through SQL injection" can then be eliminated from the attack tree as described in detail later.

The ISO/IEC 27005 asset list is predestined for this purpose because it presents a fine-grained overview of various asset classes covering all kind of possible attack targets in information security. Besides technical categories like hardware and software it also considers non-technical categories like personnel. However, particularly for domain-specific attack scenarios it makes sense to refine these assets with respect to attack-relevant characteristics. For example, the asset class smart meter (which is relevant for the electric sector) could be differentiated with respect to the supported remote data transmission standard (GSM / GPRS, WiFi, Bluetooth, Ethernet etc.). So an energy provider that does not use any smart meter supporting a WiFi transmission is not exposed to the respective attacks.

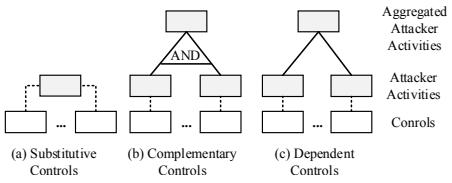


Figure 5: Modelling Rules for Controls

2.1.5. Assigning Controls

Then, security controls are assigned to the attack trees in order to get ACTrees that link the defence and the attacker perspective (see A5 in Fig. 3). ACTrees thereby enable determining to which extent the implemented security controls protect against attack scenarios and their associated adverse impacts. The starting point for assigning controls to the attacker activities is a controls list from an established standard (e. g. ISO/IEC 27002 [18]). Depending on the analysis scope a more specific standard (e. g. ISO/IEC 27019 [19] for the electric sector) can further improve the analysis results. Since the ISO/IEC 27002 standard covers roughly 110 security controls, their maturity levels can be assessed by most SMEs within a reasonable time. Although the LiSRA framework was designed with special consideration for the ISO/IEC 27000 series it also compatible with other control lists. There also exist many mappings between different control catalogues.

When assigning controls to attacker activities, the controls' relationship to each other must be considered to avoid an over-investment in security. We differentiate three relationship types: substitutive controls, complementary controls and dependent controls. The respective modelling rules, that are described in the following, are illustrated in Fig. 5.

- An example for substitutive controls for the attacker activity "password guessing" are the controls "information security awareness, education and training" (control 7.2.2) and "use of secret authentication information" (control 9.3.1). Both of them address aspects of password quality, although the latter one has a higher efficacy in this case. A set of substitutive controls is as strong as its best control because the best control takes effect. All substitutive controls are directly assigned to the correspondent attacker activity as illustrated in Fig. 5 (a).
- Complementary controls complement each other in improving the security. The highest security level can be achieved when both of them are implemented. For example, "information backup" (control 12.3.1) and "controls against malware" (control 12.2.1) complement each other in the protection against ransomware attacks. They are independent and have a multiplicative effect. Complementary controls are always linked with AND operations (to be treated multiplicative as described later) because attackers necessarily have to attack both of them for a successful attack. For this, intermediate attacker activities need to be added as shown in Fig. 5 (b).
- Dependent controls reflect a relationship type where the controls are only as good as the weakest control. An example is the relationship between a "physical security perimeter" (control 11.1.1) and an "access control policy" (control 9.1.1). A very refined and mature physical security perimeter control for instance can be useless if there is no access policy control in place, and vice versa. Dependent controls for an attacker activity are always modelled with OR operations where a rational attacker always chooses the weakest control (more details are described later). Here, intermediate nodes need to be added, too (see Fig. 5 (c)).

  There already exist lists of control dependencies for the ISO/IEC 27002 standard that can be used in the construction process of ACTrees [20].

Modelling rules for more complex relationship types like synergetic controls (that together produce an effect greater than the sum of their individual effects) are not considered here.

2.1.6. Assessing Control Efficacy

When the ACTrees are constructed they are parameterised, starting with the control efficacy (see A6 in Fig. 3). This parameter is determined for each "control to attacker activity" relation in the ACTree (see Fig. 4.

It reflects how effective a control will averagely (in the considered domain) protect against an attacker activity when it is correctly implemented.

For example, even a very mature security awareness program might be very effective in training employees to recognise phishing attacks but it might be much less effective against more specific and sophisticated attacks. This illustrates that the parameter is independent from the actual implementation level of a control. The experts assess the control efficacy based on experience and knowledge using a 3-point scale (low (L), medium (M), high (H)), which is subsequently mapped to [0,1]. *High*

is mapped to 1, *medium* to 0.67 and *low* to 0.33. A control efficacy of 0 would simply mean the control should be removed from the model. So the efficacy for an attacker activity i and a control j is $e_{ij} \in [0, 1]$.

2.1.7. Assessing Attack Costs

The same applies for the attack costs. Here, the term "attack costs" is not defined in a purely monetary sense but also in the sense of required resources. In practice, it can be a challenging task to assess the attack costs using a fine-grained scale. Therefore, the attack costs are estimated by the experts using the same 3-point scale (L/M/H) and the same mapping to a [0,1] scale that is used for the control efficacy, too (see A7 in Fig. 3). The attack costs are estimated for each attacker activity (that are defined in the leaf nodes). It is assumed that no attack can be performed for free. In the risk computation phase, the attack costs are aggregated up the tree according to the assumed attacker model. The details are described in Sect. 2.3.

2.2. Phase 2: User Input

When the framework has been set up by the domain experts users can specify their security practices and organisational characteristics.

2.2.1. Assessing Maturity Levels

The security practices are represented by the organisation's maturity levels of the security controls (MLSC). The maturity levels are used as a measure to quantify the implementation status of a security control. The higher the maturity level of a control, the higher is the chance that it is performed in an effective and secure way so that it contributes more to the organisational security. In the following, the COBIT maturity levels are used that are also defined in the ISO/IEC 15504 standard [21, 22]. Since the COBIT framework is used widespread in industry many security experts are familiar with its maturity levels and even use them in practice. For example, the information security assessment questionnaire from the German Association of the Automotive Industry (VDA) is also based on maturity levels of security controls following ISO/IEC 27002 and has a very high degree of acceptance within the German Automotive Industry [23]. So many organisations have already gathered these information. Furthermore, there also exist mappings between different control catalogues. The COBIT maturity levels are also similar to those of other prominent frameworks (e. g. NIST SP 800-30 [10], SSE-CMM (ISO/IEC 21827:2008)[24] and CMMI [25]). This also supports the reliability of the user input.

Since the ISO / IEC 27002 standard covers roughly 110 security controls their maturity levels can be assessed by most SMEs within a reasonable time. For very small organisations with less resources it can also be sufficient to concentrate on assessing entire control subcategories (34 items) or categories (14 items). Since the security controls are hierarchically structured the respective categories can easily be derived from the controls. There also exist several examples for similar high-level approaches in practice, for example Australia's framework for SMEs called "Essential Eight Maturity Model" that covers eight high-level controls [1] and the UK's Cyber Essentials scheme that focuses on five controls [2].

COBIT defines six maturity levels (from 0 to 5) that are normalised (by dividing the MLSC by 5) so that the MLSC for a control j ($MLSC_j$) $\in [0, 1]$. The maturity level assesses how mature the organisational processes of the controls are. Each maturity level can be achieved only when the level below has been achieved. The criteria for each maturity level are depicted in Tab. 2.

In larger organisations it can happen that one control has different maturity levels for different zones (e.g. in different departments). Following the weakest-link approach, the minimum maturity level for a control is chosen in this case. However, most SMEs might only rarely be affected by this. But even in this case one can easily deal with this problem by duplicating attack scenarios for another zone so that different maturity levels can be assigned to the same controls.

2.2.2. Reflecting Specific Organisational Characteristics

The optional user input described in this section is used to reflect specific organisational needs and infrastructural characteristics that have an effect on the organisational risk level (see A8 Fig. 6). Users have the option to select the organisation's domain, and to create, to adapt and/or to remove the ACTrees that are used to assess the own organisation.

a) *Selecting a Domain.* A very important way to refine the assessment results is to select the domain in which the user's organisation operates (e. g. the electric domain). Each domain-specific ACTree is associated with one or more domains so that the risk assessment only takes into account the attack scenarios that the user's organisation is exposed to. The domain can also be used to derive the attacker model. For example, for critical infrastructures one should reasonably assume attackers with many resources.

Table 2: COBIT 5 Maturity Levels [22]

Maturity Levels		Description
0	Incomplete	The control is not implemented or fails to achieve its purpose.
1	Performed	The implemented control achieves its process purpose.
2	Managed	The level 1 performed control is now implemented in a managed fashion (planned, monitored and adjusted) and its work products are appropriately established, controlled and maintained.
3	Established	The level 2 managed control is now implemented using a defined process that is capable of achieving its process outcomes.
4	Predictable	The level 3 established control now operates within defined limits to achieve its process outcomes.
5	Optimising	The level 4 predictable control is continuously improved to meet relevant current and projected business goals.

b) *Constructing New ACTrees.* The most powerful option to reflect specific organisational characteristics is to manually construct new ACTrees. For this, the ADTool [6] [3] has been modified with respect to the ACTrees used by LiSRA. After a user has constructed new ACTrees according to his organisation's needs they can be uploaded to the platform in order to individualise the risk assessment for their organisation.

c) *Manually Adapting Existing Trees.* Another option is to manually adapt the parameters or the structure of existing ACTrees. Changing default parameters makes sense if an organisation rates them differently, e. g. the impact of specific attack scenarios. Changing the tree structure makes sense if the organisation's infrastructure or processes significantly differ from the average.

d) *Disabling Trees.* Some of the existing trees might not be relevant for the user's organisation or they might become obsolet due to the construction of new trees or the adaptation of already existing trees. For this reason it is important that users can disable ACTrees so they are not considered for the assessment of their organisation.

e) *Semi-Automatic Adaptation of Trees.* Smaller organisations might struggle to individualise the ACTrees on their own. The most suitable way for those organisations is to make use of a semi-automatic adaptation of the ACTrees based on a short questionnaire. This questionnaire presents a hierarchical overview of asset classes (following ISO/IEC 27005) where the user marks the asset classes that do not exist in the considered risk assessment scope of their organisation (e. g. a smart meter supporting a remote data transmission over WiFi) [9]. LiSRA uses this information to automatically update the ACTrees by eliminating those attacks (trees) or attack steps (subtrees) targeting asset classes that do not exist in the scope. In case of OR operations only the respective subtree is eliminated, whereas for AND operations the parent node is eliminated because logically it cannot be successfully performed, too. The rationale behind the elimination is that attacks or attack steps that require the existence of certain assets cannot be performed without them. So the update is necessary to more precisely reflect the actual attack surface of the user's organisation.

Adaptations made by organisations can also be examined by domain experts in order to enable new or modified trees for other organisations, too. So there is an iterative improvement process in place ensuring a good quality.

2.3. Phase 3: Risk Computation

The detailed risk computation process is visualised in Fig. 6. The total risk is derived from scenario risks that are calculated based on both the probability of adverse impact and its severity.

2.3.1. Resolving Control Dependencies

As described in the previous section, the organisation's security measures are represented by security

[3] The ADTool is an open source software used for graphical modeling of attack-defense trees.

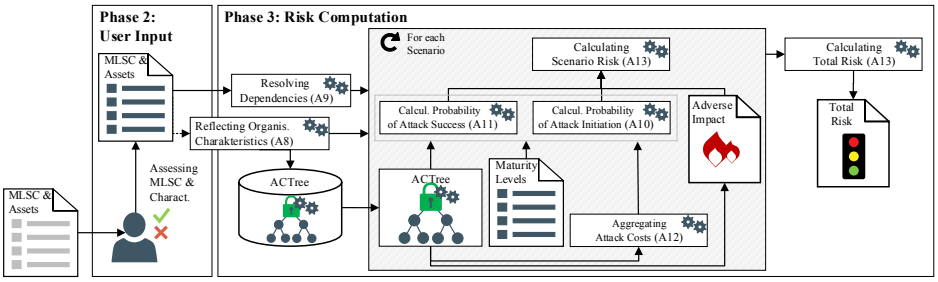


Figure 6: Phase 2 (User Input) and Phase 3 (Risk Computation)

controls following the ISO/IEC 27001. However, many of the controls are dependent on each other so that their effect cannot be assessed independently. Thus their dependencies need to be resolved (see A9 Fig. 6). If a dependent control is not mature enough it might stop other, more mature, controls from being more effective. For example, a very refined and mature physical security perimeter control can be useless if there is no access policy control in place. Sengupta systematically analysed the dependencies between all controls of the ISO/IEC 27002:2013 [20]. The results for the dependencies at the group level are visualised in Fig. 7. In the

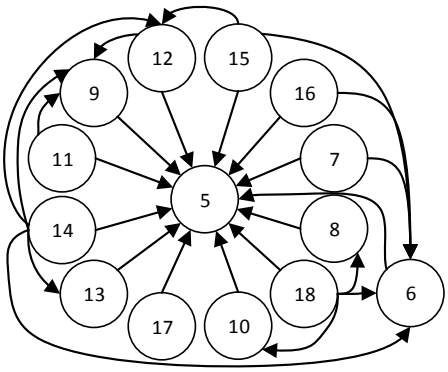


Figure 7: Visualisation of the ISO/IEC 27002 group dependencies, own figure based on [20]

following, we distinguish between strong and weak dependencies. In a strong dependency, one control strictly requires the implementation of another control. For example, the prerequisite to protect an area with a physical security perimeter (control 11.1.1) is the implementation of an access control policy (control 9.1.1). Therefore, it is a strong dependency. On the other hand, dependencies on the organisation's policy for information security (control 5.1.1) for instance, are typically weak dependencies because this policy, which should be defined and approved by the management, influences other controls to a lesser extent.

To resolve these control dependencies the dependency function $d(i)$ is applied (see Eq. (1)). Here, control i depends on the set of the controls $k \in K$. In case of a strong dependency, the MLSC of the dependent control results from the minimum MLSC of both controls. So it follows a weakest link approach. For example, a missing access control policy (MLSC=0) decreases the maturity level of the dependent physical security perimeter control to 0, even if the physical security perimeter control is implemented in a very mature way. In case of weak dependencies a control is supported by its depending controls but they are not necessarily required. So even if the other controls are not in place (MLSC=0) the dependent control can still achieve a good maturity level. This is reflected by $\Delta_{ik} = 3$ in Eq. (1).

$$d(i) = \min_{k \in K}(MLSC_i, MLSC_k + \Delta_{ik}) \tag{1}$$

with

$$\Delta_{ik} = \begin{cases} 0 & \text{if controls i and k have a strong dependency,} \\ 3 & \text{if controls i and k have a weak dependency.} \end{cases}$$

In order to resolve all dependencies the dependency function is defined recursively and is applied to all dependent controls in the ACTrees, following the dependencies identified by Sengupta [20].

2.3.2. Assessing the Probability of Attack Initiation

When the dependencies are resolved, the risk computation starts. The first step is to assess the probability of

attack initiation, $PI \in [0, 1]$ (see A10 Fig. 6). It reflects the selection probability for a specific attack option because (in case of OR operations) an attacker can choose between different attack options. We assume a rational attacker who always chooses the attack option maximising his utility. Some exemplary attacker models are defined in (Eq. 3 to Eq. 6). Here, one-shot attacks are assumed where the attacker only performs the best attack. This is modelled by the following constraint defining that the sum of the weighed decisions for a subtree is 1 (Eq. 2).

$$\sum_{j \in J} PI_{ij} = 1 \tag{2}$$

The first attacker model describes an attacker with very limited resources and a strong cost focus, e.g. script kiddies. In this case, the attacker always chooses the cheapest option.

$$PI_{ij}^{ScriptKiddie} = \begin{cases} 1 & \text{for } j = min_{j \in J} c_{ij}, \\ 0 & \text{else .} \end{cases} \tag{3}$$

The next one defines an attacker who is only interested in the attack option that maximises the probability of success, e.g. nation-state attacker. The attack decisions are not influenced by costs.

$$PI_{ij}^{Nation-StateAttacker} = \begin{cases} 1 & \text{for } j = max_{j \in J} PS_{ij}, \\ 0 & \text{else .} \end{cases} \tag{4}$$

The third attacker model represents an attacker who considers both costs and attack success and concentrates on the cost efficiency of an attack. The model determines the most cost-efficient attack decision. Since it is a good trade-off between probability of success and attack costs it might be an attacker model representing many attackers.

$$PI_{ij}^{EfficiencyMaximiser} = \begin{cases} 1 & \text{for } j = max_{j \in J} \frac{PS_{ij}}{c_{ij}}, \\ 0 & \text{else .} \end{cases} \tag{5}$$

The next model equally covers all attack options ($j \in J$) by assuming a random-shot attacker. It measure the average security.

$$PI_{ij}^{RandomShotAttacker} = \frac{1}{|J|} \tag{6}$$

Apart from those simple attacker models it is also possible to model more sophisticated ones by considering probability distributions (e.g. the standard normal distribution depending on the attacker's success chances) or by more refined utility functions, e. g. following the ideas by Ingoldsby [26].

2.3.3. Assessing the Probability of Attack Success

We define the probability of attack success, $PS \in [0, 1]$, as the probability that an attack (or an attack step), once initiated, succeeds. Thus, it is also determined by the probability of attack initiation. It is calculated using a tree-based algorithm aiming to determine to which degree the implemented security controls protect against attack scenarios or attack steps once an attack is initiated (see A11 Fig. 6).

First, the probability of attack success is calculated for the attacker activities (that are always located in the leaf nodes of the trees). The attacker's probability of success is derived by the strength of the assigned controls. It is determined by the strongest control – so a maximum function is applied (see case 3 in Eq. (7)).

$$PS_i = \begin{cases} \prod_{j \in J} PS_{ij} & \text{for inner nodes with AND,} \\ \sum_{j \in J} (PI_{ij}\, PS_{ij}) & \text{for inner nodes with OR,} \\ CF(1 - \max_{j \in J} CS_{ij}) & \text{for leaf nodes.} \end{cases} \tag{7}$$

The rationale for this is that (following the modelling rules for controls) only in case of substitutive controls more than one control can be assigned to an attacker activity; therefore only the strongest controls takes effect. Then, the probability of attack success is subsequently aggregated up the tree until the final attack goal is reached.

The control strength, $CS_i \in [0, 1]$, measures the ability of the controls j to resist against a specific attacker activity i. In general, the more mature and effective a control is the better it protects against attacks. So a control's strength is defined by the product of a control's maturity and its efficacy (Eq. (8)).

$$CS_i = \min_{j \in J}(e_{ij} \times MLSC_j, r) \tag{8}$$

The *min* function is used to model that a control strength of 1 (100 % security) can normally not be achieved. So the residual value is set to $r = 0.99$.

Since the probability of success also depends on the attacker model the control strength is weighted with a capability factor, $CF \in [0, 1]$ (see case 3 in Eq. (7)). It expresses how capable an attacker is in performing a specific attack scenario. It assumes that less capable attackers (like script kiddies) are less successful in performing complex attack scenarios than more capable attackers (like nation-state attackers), whereas they might be equally successful in performing very simple attacks.

Table 3: Attacker Capability

Attacker Model	Attacker Capability		
Nation-State Attacker	unlimited	=	∞
Average Attacker	$3 \times$ high	=	3
Script Kiddy	$1 \times$ low	=	0.33

The capability factor is defined as follows:

$$CF = min(1, \frac{AC^{attacker}}{c_s}) \tag{9}$$

The attacker's capability $AC^{attacker}$ describes how expensive an attack scenario can be for a specific attacker so that he can still effectively cope with it. These costs are not interpreted in a purely monetary sense but also in the sense of required resources which includes factors like attacker skills. Tab. 3 illustrates exemplary input values for different attacker models. Script kiddies have very limited resources and know-how so it is assumed that they might only be capable to effectively perform one attacker activity with low costs, whereas nation-state attackers potentially have unlimited resources. The quantification of cost values is the same as described in Section 2.1.7. (low=0.33; medium=0.67; high=1). NIST SP 800-30 provides additional information for quantifying attacker capabilities that can be used to further refine the input [10]. The attacker's capability is then divided by a reference value measuring the attack costs for an average attacker (like the efficiency maximiser) to execute the entire scenario. These reference value is calculated without considering the capability factor because it is only used to compare the capabilities for different attacker model with each other. These scenario costs can directly be derived from the costs for single attacker activities. More detailed information are provided in the next section (see Eq. (10).

When the weighted control strength is determined for all leaf nodes, they are aggregated up the tree in order to determine the probability of attack success for an entire attack scenario. For this, it is differentiated between inner AND nodes and inner OR nodes[4].

In case of parent nodes with AND operations the attacker does not have any choice, both attack steps have to be performed. To aggregate the probability of success all steps are multiplied by each other (see case 1 in Eq. (7). In case of parent nodes with OR nodes the attacker can choose between different attack options. For each option $j \in J$ the probability of attack success PS is weighted with the corresponding probability of attack initiation PI (see case 2 in Eq. (7). The aggregation process continues until the root node is finally reached.

[4]A parent node with only one child yields the same result a as hypothetical AND- or OR-node with one sub-node.

2.3.4. Aggregating Attack Costs

In most cases attack decisions are influenced by attack cost (e. g. for script kiddies or efficiency maximisers). So there is the need to assess the attack costs for each attack step in the attack tree (see A12 Fig. 6). For this, the initially gathered attack costs for the attacker activities are aggregated up the tree. In case of inner nodes with AND operations the attacker has to perform both attack steps so the attack costs are added up. In case of OR operations the expectation of the attack costs for a successful attack are calculated by weighting the the attack costs with the probability of initiation. So the attack costs are aggregated in the same way as the probability of attack success.

$$c_i = \begin{cases} \sum_{j \in J} c_{ij} & \text{for AND nodes,} \\ \sum_{j \in J} (PI_{ij}\ c_{ij}) & \text{for OR nodes.} \end{cases} \tag{10}$$

2.3.5. Assessing the Risk

The risk for a single scenario, $R_s \in [0, 1]$, is defined as product of the probability of attack success and the magnitude of adverse impact for a scenario s. PS_s and I_s refer to the root node of scenario s.

$$R_s = PS_s\ I_s \tag{11}$$

Finally, the total risk, $R \in [0, 1]$, adds up the weighted risk for each scenario (see A13 Fig. 6).

$$R = \sum_{s \in S} (PI_s\ R_s) \tag{12}$$

2.4. Phase 4: Recommender Application

The next step, when the risk has been computed, is to identify the most beneficial security activities.

One option is to manually inspect the results of the risk analysis. If the total risk indicates the need for action one can go through the list of scenarios to identify the high-risk scenarios. Then, users can manually inspect the respective ACTrees, e. g. to identify the most influential controls for these high-risk scenarios. A manual inspection also enables the risk assessment for very specific attack steps.

However, a faster and more objective approach for comprehensive analyses is to use the recommender application that automatises the inspection process. It can be used to get recommendations for the most effective and the most cost-efficient security activities that

are represented by MLSC increases. To further operationalise the process of improving the maturity levels, there exist mappings between the high-level ISO/IEC 27002 controls and concrete security measures (e. g. the mapping from the German Federal Office for Information Security between ISO/IEC 27002 controls and the security measures listed in the IT baseline protection [27]). Those mappings are especially helpful for MLSC increases from level 0 ("Incomplete") to 1 ("performed"). Fig. 1 visualises how these recommender application interacts with the other components. It receives the MLSC from the user and identifies beneficial security activities by simulating the corresponding risk and costs with the risk computation component.

2.4.1. Most Effective Security Activities

The first recommender application identifies the most effective security activities. It concentrates on a short-term perspective and therefore analyses the effects of incremental MLSC increases by one. The rationale for this is that improvements of organisational routines is a time-consuming process which needs to be conducted stepwise. This is also explicity pointed out in the related Capability Maturity Model (CMM) standard. They argue that skipping maturity levels is counter-productive because each level forms a necessary foundation for the next higher level which also holds for the COBIT maturity levels [28].

To identify the most effective security activities they are ranked according to the effect they have on the risk level. This measures is also known as Birnbaum measure [29]. So LiSRA increments each control's MLSC one after the other and calculates the risk reduction for each MLSC increase. Finally, all security controls with an expected risk reduction above a defined threshold are listed and sorted by the achieved risk reduction.

2.4.2. Most Cost-Efficient Security Activities

The second recommender application is based on a cost-benefit analysis and therefore relates the resulting list of the first recommender application (containing the most effective security activities) with the corresponding security costs. So cost estimations for information security costs are required for this. Here, the term "security costs" is not defined in a purely monetary sense but also in the sense of required resources.

In the following, the security costs are differentiated into the control-specific costs and the step-specific cost factor. Both of them are described below.

- The control-specific cost factor (CC) can be derived from a study by the Software Engineering Institute (SEI) in which they have empirically analysed the time needed to move up to the next MLSC. The data have been gathered with SCMAPI (Standard CMMI Appraisal Method for Process Improvement) that was conducted from 2006 to 2008 with almost 3,500 organisations. The results show that the maximum cost factor[5] for an MLSC increase is 16 [30]. This factor is reflected by the scale for control-specific cost factor depicted in Tab. 4a. For this, a geometric progression with a maximum factor of 16 and a factor to the next level of 2 is used. Brecht et al. have analysed the information security cost ratio for the ISO/IEC 27002 control categories. They can be used as rough default values[6] to estimate the security costs [31].

  However, the study refers to CMMI maturity levels that slightly differ from COBIT maturity levels in the way that COBIT level 1 ("performed") is between the CMMI's level 1 ("initial") and 2 ("managed") – it is assumed that it is exactly between level "initial" and "managed" in terms of time. The other maturity levels are basically the same [22, 25]. This has a negligible effect on the chosen cost factors.

- The security costs do not only depend on the characteristics of a specific security control but also on the concrete MLSC increase which is modelled by the step-specific cost factor (SC).

  Here, it is assumed that the time to move up from CMMI level 0 ("not performed") to 1 ("initial") is similar to the time to move up from CMMI level 1 ("initial") to level 2 ("managed").

Table 4: Costs for an MLSC Increase

(a) Control-Specific Cost Factor (CC)

Costs	Factor
Very High	4
High	2
Medium	1
Low	0.5
Very Low	0.25

(b) Step-Specific Cost Factor (SC)

Step	Factor
0→1	0.4
1→2	0.13
2→3	1
3→4	0.93
4→5	0.6

[5] The cost factor refers to the smallest and the largest observed value that is not an outlier

[6] The control categories 5,6 and 16 are associated with very high costs; category 9 with high costs; 8,11,13,14,17 and 18 with medium costs and 7 with low costs.

Accordingly, it takes 6 months to move from COBIT level 0 to 1, 2 months from level 1 to 2, 15 months from level 2 to 3, 14 months from level 3 to 4, and 9 months from level 4 to 5 [30]. This indicates how much effort MLSC improvements take and how time-consuming they are.

These effort values are now used as a weighting factor w for the security costs SC.

The step-specific cost factor is then normalised so that $SC \in [0, 1]$. Thus, an MLSC increase from 0 to 1 yields $\frac{6}{15} = 0.4$, from 1 to 2 yields $\frac{2}{15} = 0.13$, from 2 to 3 yields $\frac{15}{15} = 1$, from 3 to 4 yields $\frac{14}{15} = 0.93$, and from 4 to 5 yields $\frac{9}{15} = 0.6$. An overview is shown in Tab. 4b.

The next step is to calculate the cost efficiency CE for each MLSC increase of a control i by using Eq. (13).

$$CE_{i,MLSC} = \frac{RR_i}{CC_i \times SC_{MLSC}} \tag{13}$$

It divides the received risk reduction RR by the step-specific security costs SC that arise from an MLSC increase for control i. Then, all controls with an cost efficiency above a defined threshold are sorted and displayed. An example is shown in Sect. 4.4.

2.4.3. Providing Transparent Recommendations

Transparent recommendations are of crucial importance for the acceptance of recommender systems such as LiSRA. It describes to which extent users understand why a particular item is recommended to them [32]. Therefore, besides the recommendations themselves, also the rationale behind the recommendations is presented to the user by a graphical explanation interface.

The mitigating effects of the recommendations are presented to the user in different ways. He can choose between the scenario-centric and the recommendation-centric perspective. The scenario-centric perspective contrasts the effects of all recommendations for a specific scenario, whereas the recommendation-centric perspective illustrates the mitigating effects of a specific recommendation for each scenario. All nodes (attack steps) in the ACTrees are coloured according to the reduced probability of attack success caused by the recommended control increase. The colour coding ranges from red (no effect) to green (very high effect). The user can navigate through the trees to review the mitigating effects on each attack or attack step for each recommendation. The graphical explanation interface presents the mitigating effects of a control in the context of concrete attack steps. This serves to implement the given recommendations in a more effective way. By looking into the ACTrees decision-maker might learn that the security control "information security awareness, education & training" should be implemented with a stronger focus on phishing attacks than on other attacker activities.

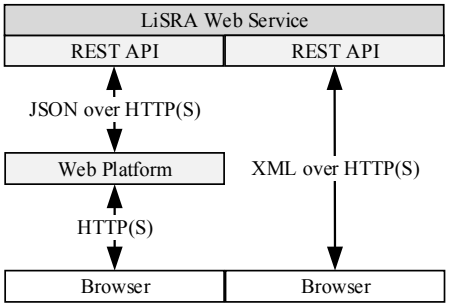


Figure 8: Architecture

2.5. Implementation

The LiSRA framework has been implemented as a RESTful web service in Java so it can easily be imported and used by other projects as well (LiSRA-as-a-Service). The high-level architecture is illustrated in Fig. 8.

The web service can for example be called over HTTP(S) with a simple browser GUI where a user uploads an XML file containing his MLSC. As return he gets back another XML document presenting the total risk as well as the specific risks for each attack scenario.

Additionally, the LiSRA framework has been integrated into the SIDATE security management web platform which has been developed in Liferay 7.0 [33]. The user enters the organisation's maturity levels in the data input section (see Fig. 9), whereupon all the risks are graphically represented in the risk representation section (see Fig. 10). For this, the web portal transmits the user's MLSC to the web service (in JSON) that returns back all the risk levels. For the sake of transparency, the corresponding ACTrees are visualised, too. The purpose of the integration was to further ease the process of going through a longer questionnaire. It aims to ease the burden of going through a longer questionnaire by enabling and motivating the user to complete or to update the MLSC along the way when interacting with other parts of the platform [8].

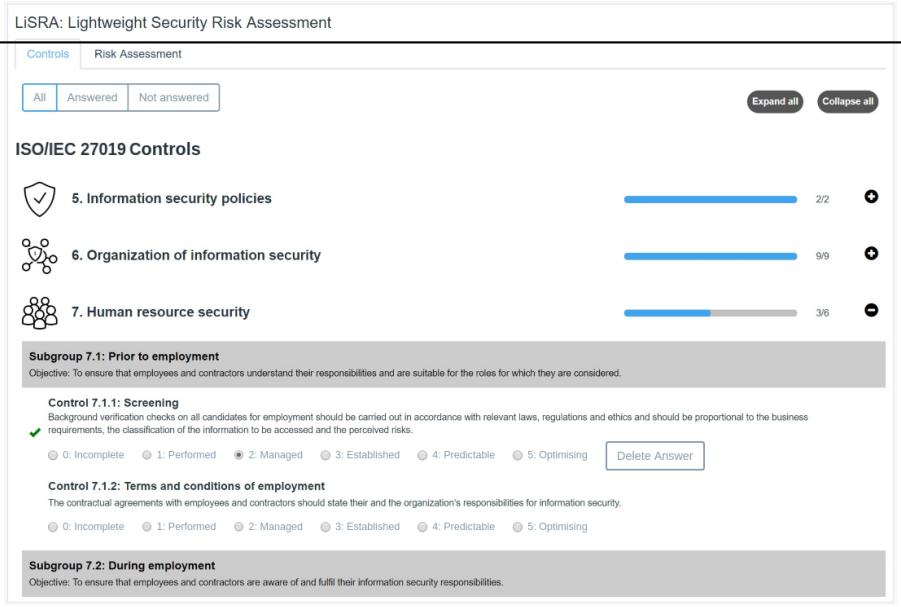


Figure 9: Data Input Section

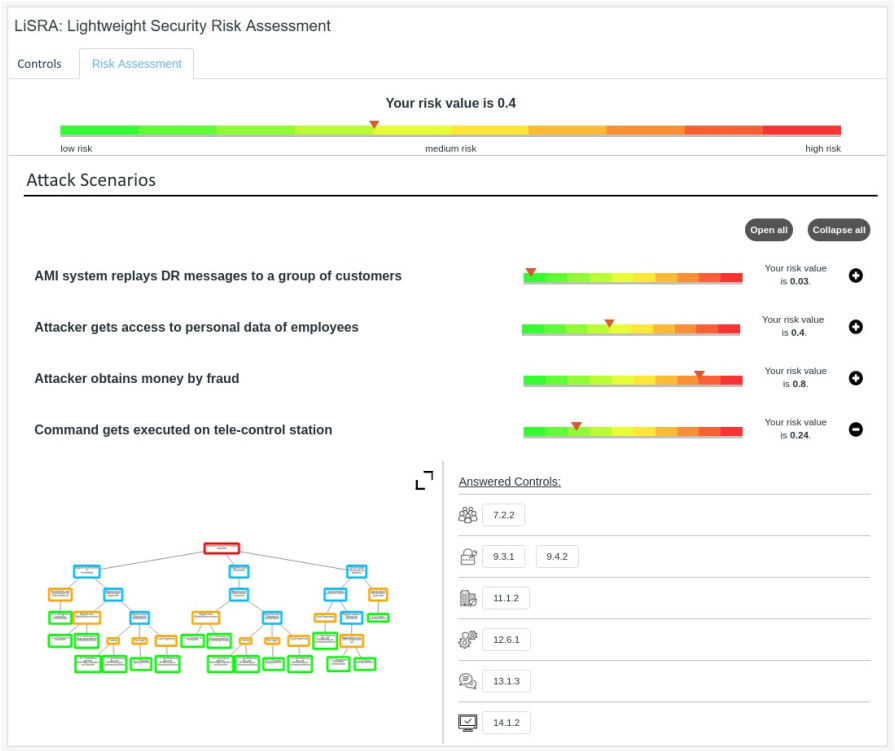


Figure 10: Risk Representation Section

3. Example

In this section we demonstrate for the exemplary domain of the electric sector that LiSRA can be used with little extra effort.

3.1. Phase 1: Expert Input

In phase 1 the experts construct and parameterise the ACTrees.

3.1.1. Identifying Attack Scenarios

The first step is to identify relevant attack scenarios. For the elctric sector there exist a well elaborated collection of attack-defense trees including the corresponding impact categories that can be used as initial input for the framework. They are provided by the National Electric Sector Cybersecurity Organization Resource (NESCOR) [15]. Although their trees are represented in a different way (so they need to be transformed), it makes much sense to use them as a starting point. Generally, it is recommended to built on already established material in order to save time and costs and to improve quality.

For the exemplary application of the model we use the simple attack scenario illustrated in Fig. 11 where the attacker tries to steal a server.

3.1.2. Assessing Adverse Impact

The impact assessment scale is illustrated in Tab. 1. The impact assessment always depends on the specific context of the scenarios (e.g. the assets at stake). For the given scenario we assume a severe adverse impact with $I_s = 1$.

To further refine the results it can make sense to use a domain-specific method. For the electric sector there is an impact scoring model proposed by the National Electric Sector Cybersecurity Organization [5] where the experts score the impact of scenarios based on 15 criteria[7]. For each criterion they can select one out of four choices. Depending on their answer, the criterion is scored with 0, 1, 3 or 9. The overall sum (between 0 and 135) reflects the scenario's impact. For the reason of s implicity, the scoring model is not used in the example.

3.1.3. Constructing Attack Trees

The ACTree used in the exemplary attack scenario (see Fig. 11) is a simplified tree only used for demonstration purposes and to explain how LiSRA works. As

[7]Examplary criteria are "negative impact on customer service", "negative impact on billing functions" or "restoration costs".

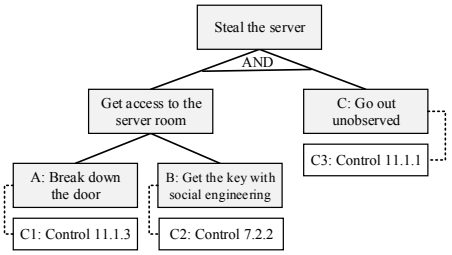


Figure 11: Exemplary Attack-Control Tree

defined in Section 3, the root node of the tree represents the attack goal of the scenario and all attacker activities are located in the leaf nodes. The presented scenario is inspired by Bistarelli et al. [16]. The attack goal is to steal a server. To achieve this, the attacker must have access to the server room and must go out unobserved (attacker activity C). There are two option to get access to the server. He can either break down the door (attacker activity A) or he can get the key using social engineering (attacker activity B).

3.1.4. Assigning Assets

In the given example the rood node is obviously associated with the general asset class "server". A high-level perspective is sufficient in this case because the attack scenario is very high-level, too.

3.1.5. Assigning Controls

When the attack trees have been constructed the corresponding security controls are assigned. As recommended above, the control list from ISO/IEC 27002 can be used respectively the more specified ISO/IEC 27019 which addresses the special needs for the electric sector.

In the given simplified example three controls are assigned. A protection against attacker activity A ("break down the door") is control C1 (11.1.3) which addresses "securing offices, rooms and facilities". The second attacker activity "get the key with social engineering" can be mitigated by control C2 (7.2.2) which is about "information security awareness, education and training". Control C3 (11.1.1) is about "physical security perimeter" which comprises for instance video surveillance. So it protects again the attacker activity of "going out unobserved"

3.1.6. Assessing Control Efficacy

Next, the ACTrees are parameterised. The control efficacy depends on the context so it is individually as-

sessed for each associated attacker activity. For example, the control "securing offices, rooms and facilities" is assumed to be effective against breaking down a door so its efficacy is assessed as "high" ($e_{C1} = high \Leftrightarrow e_{C1} = 1$), whereas the general control "awareness, education and training" is assumed to be less effective against specific social engineering attacks ($e_{C2} = medium \Leftrightarrow e_{C2} = 0.67$).

3.1.7. Assessing Attack Costs

The attack costs are gathered for each attacker activity using the 3-point scale defined in Section 3. In the present example it is assumed that the costs to get the key with social engineering are significantly higher ($c_B = high \Leftrightarrow c_B = 1$) than to break down a door ($c_A = medium \Leftrightarrow c_A = 0.67$) which is again assumed to be more expensive than going out unobserved ($c_C = low \Leftrightarrow c_C = 0.3$).

3.2. Phase 2: User Input

3.2.1. Assessing Maturity Levels

After the initialisation phase the user enters his organisation's MLSC. For control C1 it is assumed that there are established processes that are performed in the entire organisation to make sure that offices, rooms and facilities are protected. Therefore, $MLSC_{C1} = 3$. Information security awareness trainings (control C2) are irregularly performed but not in a managed way so $MLSC_{C2} = 1$. The processes addressing physical security perimeters (control C3) are systematically monitored and measured at an organisational level so $MLSC_{C3} = 4$. Finally, all maturity levels are normalised between 0 and 1 (by division by 5) so that $MLSC \in [0, 1]$.

3.2.2. Reflecting Specific Organisational Characteristics

Since it is assumed that the given organisation has servers in place (which is the only asset class associated with $Scenario_1$) the tree is fully considered in the risk assessment. Otherwise, it the entire scenario or attack steps would be excluded from the analysis.

3.3. Phase 3: Risk Computation

The risk computation process, visualised in Fig. 6, starts with resolving the control dependencies. Afterwards, the risk is computed based on attack scenarios.

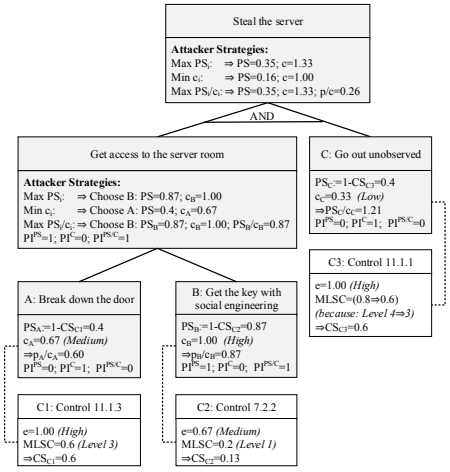


Figure 12: Exemplary Attack-Control Tree with Parameters

3.3.1. Resolving Control Dependencies

According to Sengupta's list of control dependencies, there is a strong dependency in the present example. Control C1 (11.1.3) depends on control C3 (11.1.1) [20]. Inserting their MLSC into the dependency function (Eq. 1) yields $min(3, 4) = 3$ wherefore the *effective* MLSC for control C3 is decreased by one ($MLSC_{C3} = 4 \rightarrow MLSC_{C3} = 3$). For reasons of simplicity, only the controls depicted in the ACTree are considered. Otherwise the control 11.1.2 would have to be analysed as well as the dependent controls in group 5 and 9 that are indicated in Fig. 7

3.3.2. Assessing the Probability of Attack Initiation

The probability of attack initiation reflects the selection probability for a specific attack options. In this example, we consider an attacker who always chooses the attack options with the maximum efficiency that is represented by $PI^{EfficiencyMaximiser}$.

3.3.3. Assessing the Probability of Attack Success

The probability of attack success for an attack scenario is derived from the attacker's probability of attack success for each attacker activity which is calculated by the strength of the assigned security controls. The results are also graphically illustrated in the ACTree depicted in Fig. 12. The first attacker activity A ("break down the door") is associated with one control: control C1 ("securing offices, rooms and facilities") that

Table 5: Input Parameters for Attack-Control Trees

Attacker Activity			ISO/IEC 27002 Security Controls				
ID	Description	Costs	ID	Description	Efficacy	Maturity	Strength
A	Break down the door	Medium	11.1.3	Securing offices, rooms and facilities	High	3	0.4
B	Get the key with social engineering	High	7.2.2	Information security awareness, education and training	Medium	1	0.87
C	Go out unobserved	Low	11.1.1	Physical security perimeter	Medium	$4 \rightarrow 3$	$0.2 \rightarrow 0.4$

includes for instance burglar resistant doors. So the control's efficacy for this attack is assumed to be high ($e = high \Leftrightarrow e = 1$) and the organisation's MLSC in the scenarios is 3 ($MLSC = 3 \Leftrightarrow MLSC = 3/5 = 0.6$). Then, the efficacy and the MLSC are used to calculate the controls strength ($CS_{C1} = min(0.6 \times 1), 0.99) = 0.6$.

To determine the probability of attack success the capability factor has to be assessed first. Assuming an average attacker with an attacker capability of $AC^{EfficiencyMaximiser} = 3$ (see Tab. 3) and average attack costs (for the efficiency maximiser) to perform the scenario of $c_s = 1.33$ (see Fig. 12) the capability factor yields $CF = min(1, 3/1.33) = 1$. Therefore, the probability of attack success is $PS_A := 1(1 - 0.6) = 0.4$. The same is done for attacker activity B ($PS_B := 1(1 - CS_{C2}) = 1(1 - min(0.2 \times 0.67), 0.99)) = 1(1 - 0.13 = 0.87$), and for attacker activity C (whose maturity level was decreased due to the control dependencies) ($PS_C = 1(1 - CS_{C3}) = 1(1 - min(0.6 \times 1, 0.99)) = 1(1 - 0.6) = 0.4$)

When the probability of attack success has been calculated for each attacker activity, the values for the parent nodes are calculated. The first parent node ("Get access to the server room") uses an OR operation so an attacker can decide between the attack steps A and B. The decisions is made based on the considered attacker model. In case of the efficiency maximiser (Eq. 5) activity B is chosen (because $0.87 > 0.60$). So in this case the parent node ("Get access to the server room") continues with the values for attack step B. The next parent node uses an AND operator. The attacker has to perform both attack steps so the respective probabilities are multiplied with each other. For the efficiency maximiser the probability of attack success for the scenario ("steal the server") is $PS_s = 0.87 \times 0.4 = 0.35$ and and the corresponding attack costs are $c = 1.33$.

3.3.4. Aggregating Attack Costs

The costs that are aggregated using Eq. (10) are presented in Fig. 12, following the same aggregation logic as in the previous section.

Table 6: Effects of the MLSC Increase for Control 7.2.2 (C2) from $MLSC = 1$ to $MLSC = 2$

	$Scenario_1$	$Scenario_2$	...
Prob. of Success	0.35	0.08	...
Attacker Costs	1.33	0.5	...
Attack Efficiency	0.26	0.16	...
Prob. of Initiation	1	0	...
Impact	1	1	...
Scenario Risk	0.35	0	...
Total Risk		0.35	

(a) Before MLSC Increase

	$Scenario_1$	$Scenario_2$	...
Prob. of Success	0.29	0.08	...
Attacker Costs	1.33	0.5	...
Attack Efficiency	0.22	0.16	...
Prob. of Initiation	1	0	...
Impact	1	1	...
Scenario Risk	0.29	0	...
Total Risk		0.29	

(b) After MLSC Increase

3.3.5. Assessing the Risk

Since LiSRA is a scenario-based approach the risk is first calculated for each scenario, whereupon the risk are aggregated. For a better illustration the hypothetical attack scenario 2 is added. The risk scores for $scenario_1$ and $scenario_2$ are depicted in Tab. 6. Inserting them in Eq. 11 yields $Risk_1 = 0.35 \times 1 = 0.35$ and $Risk_2 = 0.08 \times 0 = 0$. The procedure is repeated for each scenario. Finally, the organisation's total risk is calculated by adding up the weighted scenario risks according to the considered attacker model. The efficiency maximiser would choose $Scenario_1$ which has the best cost-success ratio, so the total risk is 0.35.

3.4. Phase 4: Recommender Application

The recommender application recommends the most effective and the most cost-efficient security activities in a short-term perspective.

Table 7: Simulation of Incremental MLSC Increases

	C1↑	C2↑	C3↑	...
Before MLSC Increase	3	1	4	...
After MLSC Increase	4	2	5	...
Total Risk Reduction	**0.18**	0.06	0	...
Security Costs	0.93	0.07	0.6	...
Cost Efficiency	**0.19**	0.12	0	...

3.4.1. Most Effective Security Activities

To determine the most effective security activities each control's MLSC is one after another incremented by one in order to simulate the caused risk reduction. The result is shown in Tab. 7. Improving control C3's MLSC does not cause any risk reduction because the dependency with control C1 stops C3 from being more effective. On the other hand, an MLSC increase of C1 also has a positive effect on the MLSC of C2 because C2 is not limited anymore from C1. So an increase of C1 causes the highest risk reduction.

Tab. 6 shows the effects of an MLSC increase of C2 in detail. It is assumed that C2 is not covered by the second scenario. The MLSC increase significantly reduces the probability of attack success ($PS_{S1} = 0.35 \rightarrow PS_{S1} = 0.29$) and the attack efficiency for the first scenario. The same holds for scenario risk ($R_{S1} = 0.35 \rightarrow R_{S1} = 0.29$) and for the resulting total risk ($R = 0.35 \rightarrow R = 0.29$).

3.4.2. Most Cost-Efficient Security Activities

The recommender application also identifies the most cost-efficient security activities. It takes the list with the achieved risk reduction (from most effective security activities) as basis and relates it with the arising control-specific costs CC_i and the step-specific cost factor SC_{MLSC} to reflect the MLSC increase. The simulated efficiency per MLSC increase is depicted in Tab. 7.

Low security costs are assumed for control C1 ($CC_{C1} = medium \Leftrightarrow CC_{C1} = 1$) with a step-specific cost factor for the MLSC increases from 3 to 4 of $SC_3 = 0.93$ which makes total costs of 0.93. Low security costs are assumed for C2 ($CC_{C2} = low \Leftrightarrow CC_{C2} = 0.5$) with a step-specific cost factor for MLSC increases from 1 to 2 of $SC_1 = 0.13$ which results in total costs of around 0.07. The same is done for C3 which causes costs of 0.6.

After dividing the risk reduction by the total security costs, it can be seen that an increase of C1 is the most cost-efficient security activity (see Tab. 7).

3.4.3. Providing Transparent Recommendations

In order to implement the recommended MLSC increases more effectively users can navigate through the tree and compare the mitigating effects (measured in risk reduction) for the recommended security activities. The visualisation in Fig. 13 illustrates the recommendations in a scenario-centric perspective that indicates the effects of the MLSC increases for $Scenario_1$. The graphical representation also shows the indirect effect of control C1 to the attacker activity C that is caused by a dependency. Besides that, it indicates that decision-makers should implement C1 with a special emphasis on the protection of doors (see Fig. 13a). It also highlights the importance for control C2, that normally covers very general trainings and awareness activities, to explicitly address social engineering issues (see Fig. 13c).

4. Evaluation

The evaluation of security management frameworks is a challenging task, especially because there does not exist any gold standard that could be used to conclude validity. Verendel surveyed 90 papers on quantified security where he systematically analysed which methods have been used for validation. He points out that in most cases an explicit empirical validation is missing (except for vulnerability discovery models) [34]. This is because "measuring security is hard" as Pfleeger et al. state [35]. This holds in particular for risk assessment at an organisational level because it typically deals with very complex targets of evaluation and a large scope.

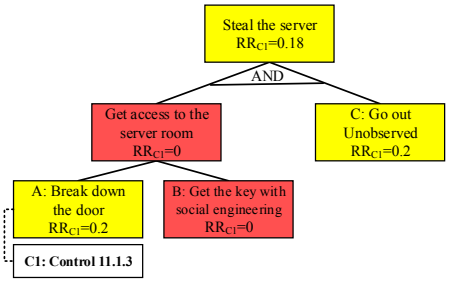
However, various important aspects of the framework have been evaluated like its applicability which has been analysed by performance tests, by analyses of robustness, and in initial qualitative evaluations. Moreover, we have examined the perceived usefulness as well as the concerns of sharing sensitive data .

4.1. Robustness

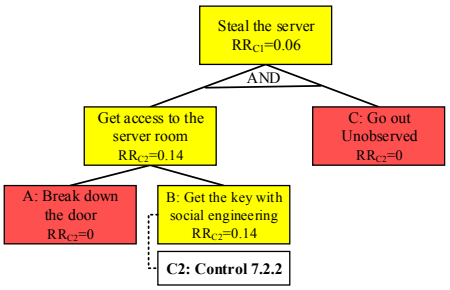
The quality of the risk assessment strongly depends on the robustness of the ACTrees. It is essential that the computed risk is robust against logical transformations (e.g. with respect to the associative or the distributive law) of the tree structures. The mathematical proofs that the computation of the probability of attack success is robust against logical transformations are presented in Appendix A.

Another aspect that is related herewith is the robustness with regard to the abstraction level of attack scenarios. It is possible to merge independent attack trees

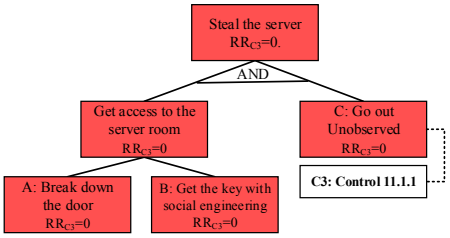
Figure 13: Visualisation of the Risk Reduction (RR) for the Recommended Controls



(a) Effects of the MLSC Increase for Control 11.1.3 (C1)



(b) Effects of the MLSC Increase for Control 7.2.2 (C2)



(c) Effects of the MLSC Increase for Control 11.1.1 (C3)

with OR operations in order to construct a larger tree at a higher abstraction level. Both equivalents must produce the same risk. This is the case because the PI (probability of initiation) function is applied to both. It does not matter if a tree is represented as a single tree or as a subtree, as long as the impact I is assessed correctly.

4.2. Performance

The practical applicability of the framework is an essential factor to be used in practice. An implementation-independent measure for this is the time complexity of an algorithm. Let i be the maximum number of attacker activities in an ACTree, then the tree consists of i leaf nodes and of maximum i inner nodes. In total, it makes a maximum of $2i$ nodes for each tree. For each of the $2i$ nodes some parameters (PI, PS, c, AC and CS) are calculated. Even though PS and c are defined recursively they only need to be calculated once for each node. The parameters are calculated with cheap operations like multiplications, additions or comparisons (for a constant set of controls) that require constant time. The nodes' parameters directly result from their child nodes. Since a node cannot have more than i child nodes the worst-case time complexity to assess a scenario's risk is $\in O(i^2)$. For all scenarios the worst-case time complexity is $\in O(s\ i^2)$. Since it is not possible that a tree has a height of i and each node has i child nodes at the same time the presented complexity analysis is very conservative so that the complexity might be even better with less strict constraints. However, the number of scenarios and the number of attacker activities in a tree are typically not very high so their risk can be assessed in a reasonably time, even for the worst-case.

The performance of the framework has also been tested with several performance tests. We have analysed the performance of the web service (for both local and remote calls) and for the web platform (for local calls). The performance tests were conducted for different numbers of realistic ACTrees (20, 50, 100, 200, 500). The ACTrees had an average number of 36.06 nodes and an average depth of 4.94 nodes. The performance tests for the web service have been automatically executed by a script logging the mean value, the median and the standard deviation (SD) for 100 service calls. The web platform has been manually tested with 5 calls using the Chromium browser version 71. The performance tests for the local web service and the web platform were conducted on a laptop with a 1.8 GHz processor and 8 GB RAM. For the remote web service tests the web service has been installed on a Tomcat server being located on a virtual private server in the same country.

It has a dual-core processor with 2 GHz and 4 GB RAM (without hyperthreading).

Tab. 8 presents the measured times which demonstrates the practical applicability of the framework from a performance perspective. The remote web service can easily handle even a vast amount of 500 attack scenarios. Applying the risk computation of phase 3 takes around 0.225 seconds (median value) for 500 scenarios. For the application scenario of recommending security activities more iterations are needed. Considering the 110 security controls of ISO/IEC 27002 and 500 ACTrees, it would approximately take 24.75 sec. However, these computations could be computed in parallel, i.e. by running several instances of the web service in parallel.

Expectedly, the page load time in a browser is significantly higher than when accessing the web service directly. The most time-consuming factors are the rendering and the scripting.

However, in practice one would expect a significantly lower number of attack scenarios. Furthermore, the source code was developed in a prototypical way without focussing on time efficiency so there is much potential to reduce the performance time.

The performance tests have also shown that the tree structure has no influence on the performance of the algorithm. This has been tested by simulations with a number of ACTrees and their transformed equivalents (n=100). The performance was directly measured in the web service. The median values were 103.4 ms and 103.9 ms.

4.3. Perceived Usefulness

The perceived usefulness of the SIDATE security management platform has been evaluated in a workshop [8]. One central part of the platform is the LiSRA framework which has been evaluated in a focus group of ten experts from eight small or medium-sized energy providers. Most of them had a profound security background and have gained experiences with ISO/IEC 27001 certification as auditor or customer.

First, a live demo of the web platform has been presented. Due to time limitations it has been focused on the conceptual ideas and is has not been gone in-depth. The attendees could interrupt at all point in time to ask any kind of questions. Afterwards, a moderated discussion was initiated where the experts were asked for general feedback and for suggestions for improvement based on their own experiences.

A central aspect of the discussion was the relevance for the ISO/IEC 27001 certification process. The experts agreed that the framework would be helpful for an internal pre-audit that takes place before the official ISO/IEC 27001 audit starts. They also emphasised that it would make a lot of sense to go through the ISO/IEC 27002 respectively 27019 controls because this would reflect what the auditor checks in the end.

In terms of suggestions for improvement they mentioned the idea to add a recommender feature that was not implemented at this time.

4.4. Concerns of Sharing Sensitive Information

For another study, the concerns of sharing sensitive data in the security management platform have been analysed, including the implemented LiSRA framework [33]. Two workshops have been conducted with experts from small and medium-sized energy providers (seven experts from six energy providers in the first workshop; six experts from five energy providers in the second workshop). The only, but sensitive, user input of the LiSRA framework are the maturity levels of the security controls. The experts did not have any concerns with sharing their maturity levels with the platform provider as long as they get a benefit out of it. Similar insights can be derived from the acceptance of the TISAX (Trusted Information Security Assessment Exchange) platform in the German automotive industry. TISAX is a sector-specific exchange platform for the German automotive industry where the results of a standardised security self-assessment (VDA-ISA) can be shared with other companies[36]. However, the data processing could also be done locally so that there would not be the need to transfer the maturity levels to an external server.

4.5. Limitations

The framework is not without limitations. First, the modelled attacker strategies only reflect one-shot attacks, that is an scenario where the attacker attempts to attack an organisation only once. He performs the best attack strategy (maximising his utility) and he does not try the second or the third-best option if he was not successful. Especially attackers with unlimited resources might follow a multiple-shot strategy.

Another limitation is that the framework is designed in particular for SMEs where a control is typically assigned to one maturity level only. In larger organisations it can happen that one control has different maturity levels in different zones. However, LiSRA can deal with this problem by duplicating attack scenarios for another zone where different maturity levels can be assigned the same controls.

Table 8: Performance tests (measured in ms)

ACTrees	Web Service (local call)			Web Service (remote call)			Page Load Time (browser)		
Quantity	Mean	Median	SD	Mean	Median	SD	Mean	Median	SD
20	95.4	93.93	5.25	137.2	134.9	6.61	5,983	5,865	294.14
50	97.81	96.27	4.30	149	144.6	24.28	7,010	7,114	450.84
100	101.28	99.83	3.63	155.5	151.4	19.39	9,505	9,516	517.9
200	110.3	107.9	6.56	192.8	190.1	12.71	12,739	12,318	819.56
500	132.8	129	7.72	227.4	224.8	18.96	26,243	25,994	701.29

5. Related Work

LiSRA is an information security risk assessment framework that also gives recommendations on future security activities. Related work for both fields of research are presented in the following.

5.1. Information Security Risk Assessment

Many literature reviews on risk assessment methodologies have been conducted in the last years [37, 38, 4, 39, 40, 41, 42]. They demonstrate that there is a lack of lightweight and reasonable frameworks that can be applied by SMEs. They provide evidence that most approaches require security-related information that are not available and that are very challenging to gather, especially for SMEs. It also becomes clear that other explicit SME approaches have far less informative value than LiSRA. An example is the model proposed by Bojanc et al. that asks for concrete values for the threat probabilities, the asset vulnerabilities and for the quantification of different loss factors [43]. It is similar for the FAIR framework that aggregates input parameters following a risk taxonomy in order to derive an asset's risk [44]. This requires the user to first define individual aggregation rules for each children-to-parent relation in the taxonomy because they strongly depend on organisational characteristics. Besides that, it is also not defined how to apply the model in order to assess the entire organisational risk. Another example is the approach by Pieters et al. that assesses the adversarial risk for an attack scenario on the basis of complex functions that are used to derive the attack success. It is very difficult to parameterise the functions, particularly for SMEs. Their approach also does not consider which security controls are in place, let alone how mature they are. On the other hand, it is one of the few models that explicitly takes into account the attacker knowledge level [45]. Karabacak et al. propose ISRAM (information security risk analysis method) – a risk assessment framework that aims to improve the quality of inaccurate input data using a survey-based method where the probability of occurrence and the consequence of occurrence are assessed for each attack scenario in two independent surveys. Although this method can improve the quality of non-available input data it still requires a sufficient number of experts with good "knowledge and awareness on the information security problem, its effects and its probable causes" [46]. They are necessary to identify and to adequately evaluate all relevant attack scenarios. So for most SMEs who typically lack in security experts it is not a suitable solution, also because of the organisational overhead that might exceed their security capacities [39].

Apart from that, many frameworks only provide extensive process descriptions and guidelines. This holds for example for OCTAVE-S [3] but also for numerous other approaches [4]. This can be challenging in particular for SMEs that usually have less capacities to become acquainted with comprehensive frameworks.

But there do exist other approaches that are designed for SMEs aiming to explicitly address their special needs. Two of the most prominent examples are Australia's framework for SMEs called "Essential Eight Maturity Model" and the UK's Cyber Essentials scheme [1, 2]. However, they only cover eight respectively five high-level security controls which makes clear that their informative value is far less than LiSRA's. The same also applies to other approaches like the analytic hierarchy process (AHP) based approach by Schmid and Pape that provide less informative value [47].

5.2. Economics of Security Activities

Since Ross Anderson argued for the importance of the economic perspective in information security in 2001 [48] and the Gordon–Loeb model raised interest in 2002 [49], extensive work has been done in the area of economic evaluation of information security activities. A literature review from 2017 on the economics of security investments systematically documents the challenges for many existing evaluation approaches [50]. It shows that many evaluation approaches for security ac-

tivities use information risk assessment approaches as a basis. So the limitations of general risk assessment approaches also apply for many evaluation approaches. So most approaches require non-available data that is hard to estimate and require in-depth knowledge in security, and can therefore not be applied by SMEs. A similar picture is also drawn in both survey paper by Neubauer [51] and by Ruan [52]. This documents that designing a lightweight framework with low requirements on the expected user input is a hard problem and still a challenging task. Good examples for this are the approach by Benaroch that expects probability distributions of investment outcomes as input data [7], and the approach by Manusco et al. where one first has to model the conditional probability tables for each scenario as basis for Bayesian networks [53].

There are also many approaches in literature that are defined very high-level. This applies for several RoSI (return on security investment) approaches that ask for high-level parameters like the annualised rate of occurrence that is challenging to estimate. This applies to the approach by Bistarelli et al. that evaluates and compares different security measures based on their return on security investment (RoSI) and their return on attack (ROA) [16]. Another common issue is that the status of high-level security controls descibing complex processes (e. g. ISO/IEC 27002 controls) is represented using a binary scale asking only for its presence [54]. This does not reflect the large spectrum of the possible implementation level at all.

Another crucial weakness of many existing approaches is that they evaluate security measures in isolation of measures already in place and that the effects of overlapping measures are often ignored by assuming they are independent from each other. They also do not reflect that different measures can be of complementary, substitutive or dependent nature which leads to an over-investment in security. This shortcoming becomes evident from broad literature reviews on security investment models [50, 51, 55]. Benaroch points out this weakness very clearly [7] referring to a number of existing work. Sawik, for example, writes that "The blocking effectiveness of each countermeasure is assumed to be independent whether or not it is used alone or together with other countermeasures" [56]. Tsalis et al. explain that "an asset is protected by multiple controls, but these may mitigate the same threats or incidents. [...] For simplicity reasons, we will assume that the controls mitigate threat independently" [57].

The same holds for the approach by Bistarelli at al. that also neglects any direct effect between different security measures, and thus implicitly assumes substitutive controls [16]. Although most attack tree approaches strictly assume complementary effects like Mancuso et al. [53, 6], others additionaly allow to model weak dependencies between measures [54]. Apart from that, there also exist more elaborated approaches aiming to precisely model the interacting effects between different security activities. However, these models typically require non-available information [7].

It is also important to consider the dependencies between security controls when identifying the most beneficial security activities. Gadyatskaya, for instance, refers to the ISO/IEC 27002 controls but neglects their dependencies when identifying the most optimal security measures [54]. This is problematic as shown by Sengupta [20].

Furthermore, most approaches do not differentiate between different attacker models. They assume an average attacker type (with average resources and average strategies) and neglect that the probability of attack success, and thus the risk, can strongly vary between different attacker types. For critical infrastructures, for instance, one should reasonably assume more powerful attackers with more resources than for other organisations. A universally applicable framework should meet this requirement. The authors are not aware of any other economic evaluation approach for security activities that enables the user to choose between different attacker models [50, 51, 55].

A major advantage of attack tree-based approaches over other methods is that they can provide detailed information why a security activity is as good or bad as it is claimed to be, and how they can be implement in the most effective manner. They are predestined for this because the intermediate results, i.e., the (reduced) probability of attack success, are calculated for each node in the tree. This makes it possible to navigate through the tree and to compare the mitigating effects of the recommended security activities in the context of concrete attack steps. However, a basic problem with attack tree-based approaches is that the quality of the assessment results strongly depend on the assumption that the underlying algorithm is robust against logical tree transformations. However, the authors are not aware of any other tree-based evaluation approach that provide evidence for this key requirement.

Although LiSRA is a universal framework that can be individualised for different domains (as shown for the electric sector) there also exist more specialised approaches addressing technical domain-specific challenges, i. e., to take into account individual client-specific security requirements in cloud computing [58].

6. Conclusion and Outlook

Assessing information security risks is one of the core duties for decision-makers in information security. In order to allocate their finite resources efficiently they need to understand the risks their organisation is exposed to. However, there is a lack of lightweight and reasonable frameworks that can be applied by SMEs. Most approaches either require too many information or their informative value is far less than LiSRA's.

Therefore, we propose LiSRA, a lightweight framework for decision support in information security. Due to the two-sided input users can focus on specifying their security practices by entering information that many organisations have already collected. These information are linked to attack paths and to the corresponding adverse impacts in order to finally assess the total risk. Apart from that, LiSRA can also be used to identify the most effective and the most cost-efficient future security activities. It provides detailed insights on their mitigating effects that also supports decision-makers in implementing the given recommendations in an effective manner. In contrast to most existing approaches, it also explicitly considers the security activities that are already implemented, and it takes into account that multiple overlapping security activities can affect each other in a complementary, substitutive or dependent way. The framework has been implemented in a prototype and its applicability has been evaluated in quantitative and qualitative analyses.

The next step is to extend the recommender application so that it identifies the optimal security activities given a limited budget. Furthermore, concrete distribution function need to be specified and empirically tested for the attacker models.

Based on the attack-control trees already constructed it is planned to conduct a case study with real-world data to evaluate how well LiSRA performs in practice and to get firsthand feedback from the organisation's experts.

Acknowledgments

This research was funded by the German Federal Ministry of Education and Research (BMBF). Grant number: 16KIS0240. We thank Leon Alexander Hermann and Ehud Cseresnyes for their contribution to the prototype implementation.

Appendix A. Proofs for Robustness

The following section contains proofs for robustness for logical transformations of the ACTree structure. The most important rules for logical transformations are the associative and the distributive law. Proofs for both are presented in the following. Proofs for more trivial rules like the commutative law are not covered here. All possible logical transformations are based on these basic transformations. It is shown that the probability of attack success for a scenario is independent from the representation of equivalent tree structures. Because the probability of attack success (PS) depends on the probability of attack initiation (PI) (see Eq. 7), for each proof it is first shown that PI is the same for different equivalent tree structures; then, the same is done for PS.

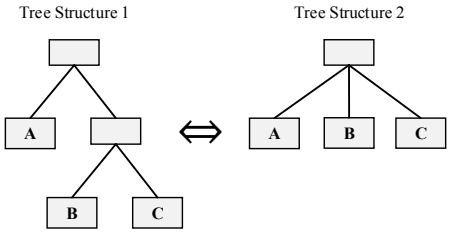


Figure A.14: Logical Transformation with Respect to the Associative Law

PI functions reflect different attacker models. They come into place in case of OR operations where an attacker can choose between different attack options. To proof the robustness for any PI function they are modelled with the generic function g. On the other hand, AND operations are modelled with function f.

The proofs also make use of the fact that, due to the logical transformations, the parameters of the nodes (here A, B and C) are the same for different equivalent tree representations.

1. **First Proof for Equivalence of Logical Tree Transformations with Respect to the Associative Law**

   First, the robustness of logical transformations is shown for the first variant of the associative law. Both equivalent tree structures are presented in Fig. A.14.

   (a) Probability of Initiation:
   According to Eq. 7, PI for tree structure 1 is represented by (A.1).

$$PI^1 = g(PI_A, g(PI_B, PI_C)) \qquad \text{(A.1)}$$

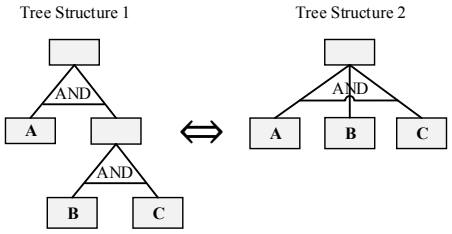


Figure A.15: Logical Transformation with Respect to the Associative Law

PI for tree structure 2 is represented by (A.2).

$$PI^2 = g(PI_A, PI_B, PI_C) \quad \text{(A.2)}$$

Therefore, assuming that function g is associative, PI^1 and PI^2 are the same for both tree structures.

(b) Probability of Attack Success:
According to Eq. 7, PS for tree structure 1 is calculated as shown in (A.3).

$$PS^1 = PI_A PS_A + \sum_{j \in J} (PI_j PS_j) \quad \text{(A.3)}$$

$$PS^1 = PI_A PS_A + PI_B PS_B + PI_C PS_C \quad \text{(A.4)}$$

PS for structure 2 is represented by (A.5).

$$PS^2 = \sum_{j \in J} (PI_j PS_j) \quad \text{(A.5)}$$

$$PS^2 = PI_A PS_A + PI_B PS_B + PI_C PS_C \quad \text{(A.6)}$$

Because $PS^1 = PS^2$, the probability of attack success is the same for both equivalent tree structures.

2. **Second Proof for Equivalence of Logical Tree Transformations with Respect to the Associative Law**

The second possible logical transformation with regard to the associative law is depicted Fig. A.15. Because the tree does not contain any OR operations the attacker does not have any attack decision. Therefore, the probability of attack initiation is 1 for all nodes.

According to Eq. 7, PS for tree structure 1 is calculated as shown in (A.7).

$$PS^1 = PS_A + \sum_{j \in J} PS_j \quad \text{(A.7)}$$

$$PS^1 = PS_A + PS_B + PS_C \quad \text{(A.8)}$$

PS for structure 2 is represented by (A.9).

$$PS^2 = \sum_{j \in J} PS_j \quad \text{(A.9)}$$

$$PS^2 = PS_A + PS_B + PS_C \quad \text{(A.10)}$$

Because $PS^1 = PS^2$, the probability of attack success is the same for both equivalent tree structures.

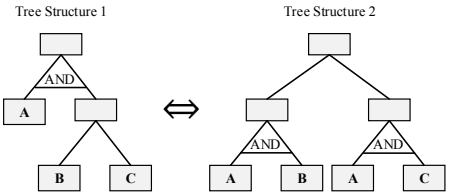


Figure A.16: Logical Transformation with Respect to the Distributive Law

3. **First Proof for Equivalence of Logical Tree Transformations with Respect to the Distributive Law**

The same is done for the distributive law. The equivalent tree structures are illustrated in Fig. A.16.

(a) Probability of Initiation:
PI for tree structure 1 is represented by (A.11) where the AND operations are modelled with function f and OR are modelled with function g.

$$PI^1 = f(PI_A, g(PI_B, PI_C)) \quad \text{(A.11)}$$

PI for structure 2 is represented by (A.12).

$$PI^2 = g(f(PI_A, PI_B), f(PI_A, PI_C)) \quad \text{(A.12)}$$

For any function g that is distributive in respect to a function f, (A.13) applies.

$$PI^2 = f(PI_A, g(PI_B, PI_C)) \quad \text{(A.13)}$$

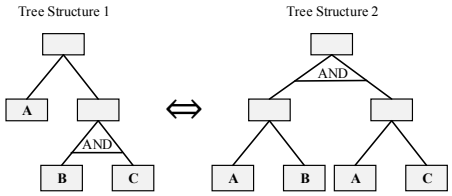


Figure A.17: Logical Transformation with Respect to the Distributive Law

Because $PI^1 = PI^2$, the probability of attack initiation is the same for both equivalent tree structures.

(b) Probability of Attack Success:
PS for tree structure 1 is calculated in (A.14).

$$PS^1 = PS_A \sum_{j \in J} (PI_j PS_j) \quad \text{(A.14)}$$

$$PS^1 = PS_A (PI_B PS_B + PI_C PS_C) \quad \text{(A.15)}$$

$$PS^1 = PI_B PS_A PS_B + PI_C PS_A PS_C \quad \text{(A.16)}$$

(A.16) can be transformed in the way that the nodes A and B resp. A and C are merged. This transformation demonstrates that PS^1 equals PS^2. The notation PI_{AB} refers to PI for the parent node of node A and B. The same holds for PS_{AB}.

$$PS^1 = PI_{AB} PS_{AB} + PI_{AC} PS_{AC} = PS^2 \quad \text{(A.17)}$$

4. **Second Proof for Equivalence of Logical Tree Transformations with Respect to the Distributive Law**

The second possible logical transformation with regard to the distributive law is depicted Fig. A.17.

(a) Probability of Initiation:
PI for tree structure 1 is represented by (A.18).

$$PI^1 = g(PI_A, f(PI_B, PI_C)) \quad \text{(A.18)}$$

PI for structure 2 is represented by (A.19).

$$PI^2 = f(g(PI_A, PI_B), g(PI_A, PI_C)) \quad \text{(A.19)}$$

For any function f that is distributive in respect to a function g, (A.20) applies.

$$PI^2 = g(PI_A, f(PI_B, PI_C)) \quad \text{(A.20)}$$

Because $PI^1 = PI^2$, the probability of attack initiation is the same for both equivalent tree structures.

(b) Probability of Attack Success:
PS for tree structure 1 is calculated in (A.21) and is transformed into (A.23).

$$PS^1 = PI_A PS_A + PI_{BC} \sum_{j \in J} PS_j \quad \text{(A.21)}$$

$$PS^1 = PI_A PS_A + PI_B PS_B \times PI_C PS_C \quad \text{(A.22)}$$

$$PS^1 = PI_A PS_A + PI_{BC} (PS_B PS_C) \quad \text{(A.23)}$$

PS for tree structure 2 is represented by (A.24) and is transformed into (A.26)

$$PS^2 = (PI_A PS_A + PI_B PS_B)(PI_A PS_A + PI_C PS_C) \quad \text{(A.24)}$$

$$PS^2 = PI_A PS_A \times PI_A PS_A + PI_A PS_A \times PI_B PS_B + PI_A PS_A \times PI_C PS_C + PI_B PS_B \times PI_C PS_C \quad \text{(A.25)}$$

$$PS^2 = PI_A PS_A (PI_A PS_A + PI_B PS_B + PI_C PS_C) + PI_B PS_B \times PI_C PS_C \quad \text{(A.26)}$$

The present equations represent logical statements. Therefore, (A.26) can be simplified into (A.27).

$$PS^2 = PI_A PS_A + PI_B PS_B \times PI_C PS_C \quad \text{(A.27)}$$

Because $PS^1 = PS^2$, the probability of attack success is the same for both equivalent tree structures.

References

[1] Australien Cyber Security Centre (ACSC), Essential eight maturity model, https://www.cyber.gov.au/publications/essential-eight-maturity-model, 2019.
[2] National Cyber Security Centre, Uk's cyber essentials scheme, https://www.cyberessentials.ncsc.gov.uk/, 2019.
[3] C. Alberts, A. Dorofee, J. Stevens, C. Woody, OCTAVE-S implementation guide, version 1.0, Pittsburgh, PA, Carnegie Mellon University (2005).
[4] A. Shameli-Sendi, R. Aghababaei-Barzegar, M. Cheriet, Taxonomy of information security risk assessment (isra), Comput. Secur. 57 (2016) 14–30.
[5] National Electric Sector Cybersecurity Organization Resource (NESCOR), Electric Sector Failure Scenarios and Impact Analyses, Technical Report, 2013.

[6] B. Kordy, P. Kordy, S. Mauw, P. Schweitzer, Adtool: Security analysis with attack–defense trees, in: K. Joshi, M. Siegle, M. Stoelinga, P. R. D'Argenio (Eds.), Quantitative Evaluation of Systems, Springer Berlin Heidelberg, Berlin, Heidelberg, 2013, pp. 173–176.

[7] M. Benaroch, Real options models for proactive uncertainty-reducing mitigations and applications in cybersecurity investment decision making, Information Systems Research 29 (2018) 315–340.

[8] C. Schmitz, A. Sekula, S. Pape, V. Pipek, K. Rannenberg, Easing the burden of security self-assessments, in: 12th International Symposium on Human Aspects of Information Security & Assurance, HAISA 2018 ,Dundee, Scotland, August 29-31, 2018, Proceedings.

[9] International Organization for Standardization (ISO), International Electrotechnical Commission (IEC), ISO/IEC 27005 Information technology – Security techniques – Information security risk management, Technical Report, 2008.

[10] National Institute for Standards and Technology (NIST), 800-30 Rev. 1: Guide for Conducting Risk Assessments, Technical Report, 2012.

[11] B. Schneier, Attack trees, Dr. Dobbs journal 24 (1999) 21–29.

[12] S. Mauw, M. Oostdijk, Foundations of attack trees, in: International Conference on Information Security and Cryptology, Springer, pp. 186–198.

[13] B. Kordy, L. Piètre-Cambacédès, P. Schweitzer, DAG-based attack and defense modeling: Don't miss the forest for the attack trees, Computer Science Review 13-14bb (2014) 1–38.

[14] M. Davarynejad, M. Ford, D. Hadziosmanovic, O. Gadyatskaya, R. Hansen, D. Ionita, H. Jonkers, A. Lenin, Z. Lukszo, S. Mauw, B. Othman, W. Pieters, C. Probst, A. Tanner, R. Trujillo, J. van den Berg, J. Willemson, Technology-supported Risk Estimation by Predictive Assessment of Socio-technical Security, Best practices for model creation and sharing (Deliverable D5.3.2), Technical Report, 2015.

[15] National Electric Sector Cybersecurity Organization Resource (NESCOR), Analysis of Selected Electric Sector High Risk Failure Scenarios, Technical Report, 2013.

[16] S. Bistarelli, F. Fioravanti, P. Peretti, Defense trees for economic evaluation of security investments, in: Availability, Reliability and Security, 2006. ARES 2006. The First International Conference on, pp. 8 pp.–.

[17] A. Roy, D. S. Kim, K. S. Trivedi, Cyber security analysis using attack countermeasure trees, in: F. T. Sheldon, S. J. Prowell, R. K. Abercrombie, A. W. Krings (Eds.), Proceedings of the 6th Cyber Security and Information Intelligence Research Workshop, CSIIRW 2010, Oak Ridge, TN, USA, April 21-23, 2010, ACM, 2010, p. 28.

[18] International Organization for Standardization (ISO), International Electrotechnical Commission (IEC), ISO/IEC 27002:2013, information technology – security techniques – code of practice for information security controls (2013).

[19] International Organization for Standardization (ISO), International Electrotechnical Commission (IEC), ISO/IEC 27019:2017, information technology – security techniques – information security controls for the energy utility industry (2017).

[20] A. Sengupta, Modeling dependencies of iso/iec 27002:2013 security controls, in: J. H. Abawajy, S. Mukherjea, S. M. Thampi, A. Ruiz-Martínez (Eds.), Security in Computing and Communications, Springer International Publishing, Cham, 2015, pp. 354–367.

[21] International Organization for Standardization (ISO), International Electrotechnical Commission (IEC), ISO/IEC 15504-5:2012, information technology – process assessment - part 5: An exemplar software life cycle process assessment model, 2012.

[22] Information Systems Audit and Control Association (ISACA), CobiT 5: A Business Framework for the Governance and Management of Enterprise IT, Rolling Meadows, 2012.

[23] Verband der Automobilindustrie (VDA), Information security assessment, https://www.vda.de/de/services/Publikationen/information-security-assessment.html, 2019, Accessed 26 February 2019.

[24] International Organization for Standardization (ISO), International Electrotechnical Commission (IEC), ISO/IEC 271827:2008, information technology – systems security engineering - maturity model (sse-cmm) (2013).

[25] M. B. Chrissis, M. Konrad, S. Shrum, CMMI for Development: Guidelines for Process Integration and Product Improvement, Addison-Wesley Professional, 3rd edition, 2011.

[26] T. R. Ingoldsby, Attack tree-based threat risk analysis (2013).

[27] G. F. O. for Information Security (BSI), Zuordnungstabelle ISO 27001 sowie ISO 27002 und IT-Grundschutz, Technical Report, 2011.

[28] M. C. Paulk, B. Curtis, M. B. Chrissis, C. V. Weber, Capability maturity model, version 1.1, IEEE software 10 (1993) 18–27.

[29] Z. W. Birnbaum, On the importance of different components in a multicomponent system, Technical Report, Washington Univ Seattle Lab of Statistical Research, 1968.

[30] Carnegie Mellon University - Software Engineering Instituteee, Process maturity profile cmmi for development scampi class a appraisal results - 2008 end-year update, Presentation, 2009.

[31] M. Brecht, T. Nowey, A Closer Look at Information Security Costs, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 3–24.

[32] P. Pu, L. Chen, R. Hu, A user-centric evaluation framework for recommender systems, in: Proceedings of the fifth ACM conference on Recommender systems, ACM, pp. 157–164.

[33] J. Dax, B. Ley, S. Pape, C. Schmitz, V. Pipek, K. Rannenberg, Elicitation of requirements for an inter-organizational platform to support security management decisions, in: 10th International Symposium on Human Aspects of Information Security & Assurance, HAISA 2016 ,Frankfurt, Germany, July 19-21, 2016, Proceedings.

[34] V. Verendel, Quantified security is a weak hypothesis: A critical survey of results and assumptions, in: Proceedings of the 2009 Workshop on New Security Paradigms Workshop, NSPW '09, ACM, New York, NY, USA, 2009, pp. 37–50.

[35] S. Pfleeger, R. Cunningham, Why measuring security is hard, IEEE Security Privacy 8 (2010) 46–54.

[36] ENX Association, Trusted information security assessment exchange (tisax), http://enx.com/tisax/tisax-en.html, 2019, Accessed 26 February 2019.

[37] European Union Agency for. Network and Information Security (ENISA), Risk Management: Implementation principles and Inventories for Risk Management/Risk Assessment methods and tools, Technical Report, 2006.

[38] S. Fenz, J. Heurix, T. Neubauer, F. Pechstein, Current challenges in information security risk management, Information Management & Computer Security 22 (2014) 410–430.

[39] A. Behnia, R. A. Rashid, J. A. Chaudhry, A survey of information security risk analysis methods, SmartCR 2 (2012) 79–94.

[40] D. Ionita, Current established risk assessment methodologies and tools, Master's thesis, University of Twente, 2013.

[41] S. M. Sulaman, K. Weyns, M. Höst, A review of research on risk analysis methods for it systems, in: Proceedings of the 17th International Conference on Evaluation and Assessment in Software Engineering, EASE '13, ACM, New York, NY, USA, 2013, pp. 86–96.

[42] C. Harpes, G. Schaff, M. Martins, B. Kordy, R. Trujillo, D. Ionita, Technology-supported Risk Estimation by Predictive Assessment of Socio-technical Security, Currently established risk-assessment methods (Deliverable D5.2.1), Technical Report, 2014.

[43] R. Bojanc, B. Jerman-Blažič, A quantitative model for information-security risk management, Engineering management journal 25 (2013) 25–37.

[44] J. Jones, FAIR - ISO/IEC 27005 Cookbook, Technical Report, The Open Group, 2010.

[45] W. Pieters, M. Davarynejad, Calculating adversarial risk from attack trees: Control strength and probabilistic attackers, in: Data Privacy Management, Autonomous Spontaneous Security, and Security Assurance, Springer, 2015, pp. 201–215.

[46] B. Karabacak, I. Sogukpinar, Isram: information security risk analysis method, Computers & Security 24 (2005) 147–159.

[47] M. Schmid, S. Pape, A structured comparison of the corporate information security maturity level, in: G. Dhillon, F. Karlsson, K. Hedström, A. Zúquete (Eds.), ICT Systems Security and Privacy Protection, Springer International Publishing, Cham, 2019, pp. 223–237.

[48] R. Anderson, Why information security is hard-an economic perspective, in: Computer security applications conference, 2001. acsac 2001. proceedings 17th annual, IEEE, pp. 358–365.

[49] L. A. Gordon, M. P. Loeb, The economics of information security investment, ACM Trans. Inf. Syst. Secur. 5 (2002) 438–457.

[50] D. Schatz, R. Bashroush, Economic valuation for information security investment: a systematic literature review, Information Systems Frontiers 19 (2017) 1205–1228.

[51] T. Neubauer, C. Hartl, On the singularity of valuating it security investments, in: Proceedings of the 2009 Eigth IEEE/ACIS International Conference on Computer and Information Science, ICIS '09, IEEE Computer Society, Washington, DC, USA, 2009, pp. 549–556.

[52] K. Ruan, Introducing cybernomics: A unifying economic framework for measuring cyber risk, Computers & Security 65 (2017) 77–89.

[53] A. Mancuso, P. Zebrowski, A. C. Vieira, Risk-based selection of mitigation strategies for cybersecurity of electric power systems, IEEE TRANSACTIONS ON DEPENDABLE AND SECURE COMPUTING (2019).

[54] O. Gadyatskaya, C. Harpes, S. Mauw, C. Muller, S. Muller, Bridging two worlds: reconciling practical risk assessment methodologies with theory of attack trees, in: International Workshop on Graphical Models for Security, Springer, pp. 80–93.

[55] L. Demetz, D. Bachlechner, To invest or not to invest? assessing the economic viability of a policy and security configuration management tool, in: The Economics of Information Security and Privacy, 2013, pp. 25–47.

[56] T. Sawik, Selection of optimal countermeasure portfolio in it security planning, Decis. Support Syst. 55 (2013) 156–164.

[57] N. Tsalis, M. Theoharidou, D. Gritzalis, Return on security investment for cloud platforms, in: Proceedings of the 2013 IEEE International Conference on Cloud Computing Technology and Science - Volume 02, CLOUDCOM '13, IEEE Computer Society, Washington, DC, USA, 2013, pp. 132–137.

[58] A. M. Nhlabatsi, J. B. Hong, D. S. D. Kim, R. Fernandez, A. Hussein, N. Fetais, K. M. Khan, Threat-specific security risk evaluation in the cloud, IEEE Transactions on Cloud Computing (2018).

B.10 On the use of Information Security Management Systems by German Energy Providers

Sebastian Pape, Christopher Schmitz, Dennis-Kenji Kipker, and Andre Sekula. On the use of information security management systems by german energy providers. Accepted for presentation at the Fourteenth IFIP Working Group 11.10 International Conference on Critical Infrastructure Protection, 03 2020

The Introduction of Information Security Management Systems by German Energy Providers

Sebastian Pape*, Goethe University Frankfurt

Christopher Schmitz, Goethe University Frankfurt

Dennis-Kenji Kipker, University of Bremen

André Sekulla, University of Siegen

Abstract

Along with other requirements, the German critical infrastructure programme required critical infrastructure providers, i.e. energy providers to implement an ISMS. We used the unique opportunity to observe the implementation and surveyed all German energy providers in autumn 2016 and 2018. Our study shows, that most of the energy providers implemented an ISMS between our surveys and reported an perceived increase in information security suggesting that the critical infrastructure programme fulfilled its purpose.

*sebastian.pape@m-chair.de

1 Introduction

Critical infrastructures are of vital importance to a nation's society and economy because their failure would result in sustained supply shortages causing a significant disruption of public safety and security. In 2016, malicious software in nuclear power plants was reported [1] followed by further reports[2][3], e.g. warnings about hackers attacking German energy providers in 2018.

With the *European Programme for Critical Infrastructure Protection* (EPCIP) and its counterpart, the German critical infrastructure protection programme KRITIS [1] governments aimed to provide the ground for more secure critical infrastructures. The new regulation challenged critical infrastructure provider in many ways. Besides general challenges such as understanding the definitions and requirements (cf. [2, p. 150ff]), and challenges from other areas, i.e. coping with the energy transition, energy providers needed to register a contact point, establish processes to report security incidents, implement security requirements following a security catalogue (§11 Abs. 1a respectively 1b EnWG), and establish and certify an information security management system (ISMS). Our investigation focuses on the introduction of an ISMS by German energy providers. For that purpose, we surveyed German energy providers in autumn 2016 when they had just learned about the requirements and in autumn 2018, roughly half a year after they had to provide the certification of their ISMS. The new regulation offers us the chance to have a

[1] German Newspaper: Spiegel Online (2016): „Schadsoftware in bayerischem Atomkraftwerk entdeckt"

[2] German newspaper: Süddeutsche Zeitung (2018): „Warnung vor Hackerangriffen auf deutsche Energieversorger"

[3] German newspaper: Süddeutsche Zeitung (2018): „Hacker haben deutschen Energieversorger angegriffen"

closer look at a large amount of energy providers introducing an ISMS to get ready for certification at the same time. We intend to investigate how the introduction of the ISMS went and how the energy providers plan to operate it. Since the real security level can not easily be measured within the survey, we are furthermore looking for evidence if the need to establish an ISMS changed the energy providers' view on security.

The remainder of this paper is as follows. Section 2 discusses the legal background of the European and German infrastructure protection programme and discusses related work. Section 3 introduces the methodology of the study and Section 4 presents the results which are discussed in Section 5. Section 6 concludes our work. Both surveys can be found in the appendix.

2 Background

2.1 European and German Political Strategies for Critical Infrastructure Protection

At an early stage, the increasing challenges of information technology protection of critical infrastructures were addressed in terms of legal policy both in the European Union and in Germany. First, in 2006 the European Union adopted the "European Programme for Critical Infrastructure Protection" (EPCIP) - also understood as a blueprint for future legislation in this area[4]. The primary aim is to protect critical infrastructures against terrorist threats. The measures proposed in the EPCIP are based on the principles of the rule

[4]EPCIP, COM (2006), 786 final, p. 3.

of law and the principle of subsidiarity enshrined in the EU, so that the measures planned by the European Commission relate less to national or regional measures and more to those of pan-European significance. Measures taken to protect critical infrastructures must also be proportionate. This means that risk and threat must be in proportion to each other. EPCIP also describes a sector-specific approach to implementing security measures. Critical infrastructures themselves are not yet defined in EPCIP; the document is rather a catalogue of measures and political guidelines for action. The framework which is proposed by EPCIP consists of several measures:

- A common procedure for the identification and designation of European Critical Infrastructures (ECI) by the way of a European Directive
- Critical Infrastructure Protection (CIP) information exchange: establishment of an EPCIP action plan, a CIP Contact Group as strategic coordinating tool, a Critical Infrastructure Warning Network (CIWIN), the foundation and use of CIP expert groups at EU level, as well as an information sharing process and the identification and analysis of interdependencies
- Contingency planning and external measures/dimensions

Also at the level of the EU Member States, policies specifically for the protection of critical infrastructure have been pursued for several years. For Germany, the "National Strategy for Critical Infrastructure Protection" (KRITIS Strategy) should be mentioned at this point, which was previously supplemented by the "National Plan for the Protection of Information Infrastructures"

(NPSI) and is now supplemented by the German cyber security strategies from 2013 and 2016. Based on the KRITIS Strategy, critical infrastructures are organisations and institutions of major importance to the state community, whose failure or impairment would result in lasting supply problems, significant disruptions to public security or other dramatic consequences[5]. In the following, further infrastructures and processing areas are listed that are critical in the aforementioned overall social sense. It should be noted, however, that these classifications are not yet legally binding, as they are only part of a political strategy:

- Basic technical infrastructures: energy supply, information and communication technology, transport and traffic, water supply and sewage disposal
- Socio-economic service infrastructures: health care, nutrition, emergency and rescue services, civil protection, parliament, government, public administration, justice, finance and insurance, media, culture

The CRITIS Strategy divides the risks and threats to such infrastructure into three categories: harmful natural events, technical and human failure, terrorism/crime and war. Based on the above-mentioned hazard situations, the strategic objectives for the protection of critical infrastructures are proposed. The focus of all government measures is on prevention and sustainability, as well as readiness to respond to serious cyber incidents. In order to achieve the objectives proposed by the KRITIS Strategy, the introduction of business continuity management and cooperation between the state and the private

[5]Nationale Strategie zum Schutz Kritischer Infrastrukturen, S. 3.

sector in the sense of a public-private partnership are addressed as a priority. In addition, the Federal Government plans to intensify international cooperation on cyber security.

2.2 European and German Legislation on Critical Infrastructure Protection

Based primarily on the European and German political strategies for the protection of critical infrastructures, various laws have been passed in recent years, which also means that there is no uniform law for the implementation of cyber security. This is also a challenge for those companies addressed by the laws. As far as IT security-specific legislation in the EU and in Germany is concerned, the following legal sources can currently be used as key drivers of corporate information security:

- the EU Network and Information Security Directive from 2016 (EU NIS Directive)
- the EU Cybersecurity Regulation from 2019 (EU CSA)
- the German IT Security Act 2015 (IT-SiG) including the BSI Critical Infrastructure Ordinance (BSI-KritisV, published in two stages in 2016 and 2017 respectively)
- the draft version of the 2nd German IT Security Act (IT-SiG 2.0, 2019)

The IT security-specific regulations, which were established by the German IT-SiG, essentially address the operators of critical infrastructures. These

are generally legally defined in §2 para. 10 BSIG and are concretized by the numerical specifications of the BSI-KritisV (so-called "threshold values"). The criteria of quality and quantity are decisive. This means that an institution is classified as critical infrastructure within the meaning of the Act if it belongs to the energy, information technology, telecommunications, transport, traffic, health, water, food, finance and insurance sectors - in this respect it is similar to the NPSI, but not congruent. In addition, in the sense of a "fault consequence relevance" as a quantitative criterion, it must be added that the infrastructure is of great importance for the functioning of the community because its failure or impairment would lead to considerable problems in the supply chain or threats to public safety. The measure of the significance of the consequences of such failures is primarily based on the figures/numbers defined in the BSI-KritisV. The German IT-SiG is a so-called "Article Law" and contains a regulatory mandate to the legislator to amend various individual laws. These include the Atomic Energy Act, the Act on the Federal Office for Information Security (BSI), the Energy Industry Act, the Telecommunications Act and the Telemedia Act of Germany. All these regulations contain special requirements for information security, which must be provided by the respective operators. In case of non-compliance, the requirements are subject to sometimes substantial sanctions. The laws themselves do not usually go into the technical-organizational details of the concrete obligations with regard to content. Thus, in most cases only general objectives to be applied to information security are defined, or reference is made to "appropriate" measures that correspond to the "state of the art". This is a so-called "undefined legal term" or a "general clause". From a legal point of

view, the "state of the art" is to be classified in the triad of "generally accepted rules of technology" and "state of science and technology", whereby the "state of the art" represents the technical-organisational mean value between these two extremes. Consequently, it depends on what is technically necessary, suitable, appropriate and avoidable in terms of malfunctions and risks at the respective present time. In addition to the technical-organisational IT security obligations in accordance with the "state of the art", critical infrastructures are also subject to a reporting obligation to the BSI. Since the IT Security Act came into force in 2015, there has been considerable speculation and uncertainty on the part of operators of infrastructures affected by the Act regarding the content and scope of the technical and organisational measures to be taken for cyber security. In the meantime, two specific guidelines have been created for the energy sector in particular to define the legal requirements, but these are outside the scope of the law itself:

- Industry-specific safety standards (B3S) Energy, based on §8a para. 2 BSIG: one standard for plants or systems for the control/bundling of electrical power (B3S Aggregators)[6] and one standard for the distribution of district heating (B3S Vv Fw)[7].

- IT security catalogue in accordance with §11 para. 1a EnWG of the supervisory authority Bundesnetzagentur (BNetzA)[8]

Both sources contain detailed specifications for the technical-organizational

[6] https://www.bdew.de/energie/b3s-aggregatoren/

[7] https://www.bdew.de/energie/b3s-fernwaermenetze/

[8] https://www.bundesnetzagentur.de/SharedDocs/Downloads/DE/Sachgebiete/Energie/Unternehmen_Institutionen/Versorgungssicherheit/IT_Sicherheit/IT_Sicherheitskatalog_08-2015.pdf

implementation of cyber security measures by the operators of energy supply networks and energy facilities, which are essentially linked to the introduction of an Information Security Management System (ISMS). The developments around a specifically European and German law of information security are finally supplemented by the EU NIS-RL, the EU CSA as well as by the draft for an IT-SiG 2.0. The NIS Directive contains obligations for so-called "essential services", which for Germany correspond to critical infrastructures. As an EU Directive, it does not have any direct effect in the Member States, but must be incorporated into German law by means of a national implementation law in order to be effective. This has already been done in 2017. In addition, legislators are increasingly creating cross-sectoral IT security-related regulations that go beyond the scope of critical infrastructures - a development that is particularly evident in the CSA and the draft of the IT-SiG 2.0. The CSA is developing a comprehensive, cross-sectoral IT security certification system that is currently still voluntary and theoretically ranges from IoT consumer products to the protection of a critical energy infrastructure. Although the IT-SiG 2.0 also introduces regulatory proposals aimed at the consumer sector, it also increases the requirements for the operation of a critical infrastructure in Germany. Among other aspects, the draft law requires that manufacturers which install their products in control systems of a critical infrastructure ensure that cyber security is guaranteed for the entire supply chain of their product. The IT-SiG 2.0 is expected to be passed by the German Parliament before the end of 2020.

2.3 Related Work

Hurst et al. [3] discuss critical infrastructures and the digital threats they face by surveying different infrastructure security strategies. Rehbohm et al. [4] did an interview study among the chief information security officers (CISOs) of the federal states of Germany about current challenges in cybersecurity management. The Federal Office for Information Security[9] (BSI) lists the status of the implementation of cybersecurity in the energy sector in 2015 [2, p. 16ff]. They state that while some of the companies have put IT security measure in place to ensure a high degree of security, other hardly have any measures in place.

Closest to our work is a study from Müller et al. [5] which also investigates ISMS for German energy provider. They called about 200 Chief Information Security Officers (CISOs) from German energy providers and ended up with 42 complete questionnaires.

3 Methodology

We surveyed German energy providers about their information security in 2016 and 2018. Besides the survey, we also got some insights by workshops within the SIDATE project [6] which showed to be useful for the discussion of the results. The SIDATE project aimed to support small and medium energy providers to cope with the security requirements. Personnel from energy providers responsible for IT security participated in the workshops p [7, 8].

[9] German: Bundesamt für Sicherheit in der Informationstechnik

3.1 Questionnaire

The questionnaire covered sections about general information, organisational aspects, ISMS and ISMS maintanance (only in 2018), the office IT, and networking and organisational aspects about the industrial control system of the energy providers [9, 10, 11]. We did pre-tests within the universities' research groups and in the SIDATE project which included project partners with domain specific knowledge. The two different versions of the questionnaire are shown in the appendix.

3.2 Data Collection

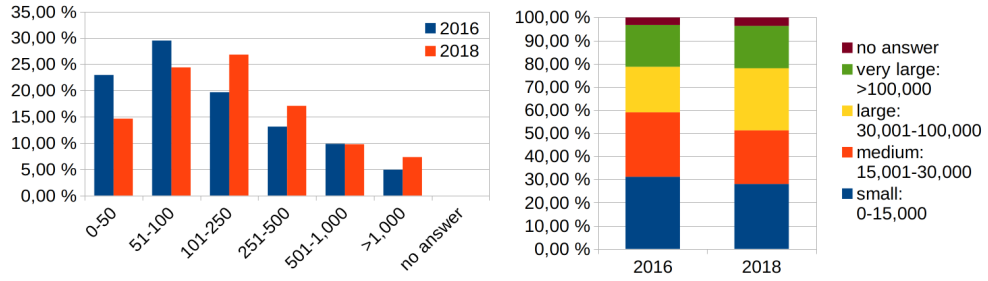
In 2016 (2018), we (physically) mailed to all 881 (890) energy providers listed in August 2016 (September 2018) [12] by the Federal Network Agency (German: Bundesnetzagentur or BNetzA), the German regulatory office for electricity, gas, telecommunications, post and railway markets [13]. We sent them a printed version of the survey and a link to the online survey along with a cover letter referring to the SIDATE project [6] about supporting small and medium energy providers with their IT-Security.

The survey lasted from September 1^{st} to October 15^{th}, 2016 (September 10^{th} to October 30^{th} 2018). and received 22 (38) replies online and 39 (46) replies by mail summing up to a total of 61 (84) replies resulting in a response rate of 6.9% (9,4%).

Since two respondents within the 2018 survey claimed that they are not regarded as critical infrastructure and therefore have not implemented an ISMS, we removed there answers.

3.3 Demographics

We asked the energy providers about the number of supply points and the number of employees as shown in Fig. 1.

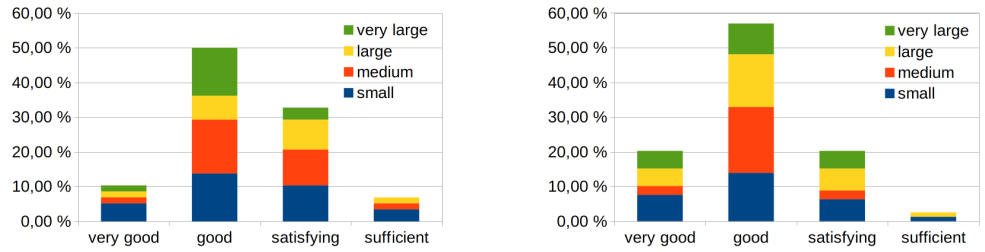


(a) Number of Employees (b) Number of Supply Points

Figure 1: Size of the participating energy providers

In order to refer to the size of the energy providers, we mapped them to the four categories "small, medium, large and very large" according to the number of supply points. In the survey, we had more distinct categories at the border (<1,000 and 100,001 - 500,000), but due to their low population we merged them. We checked with Spearman's rank correlation for similarities with the number of employees and found for 2016 (2018) ρ-values of 0.725 (0.496) with p-values lower than 10^{-5} indicating a strong (moderate) relationship. Therefore, we argue that it is sufficient to consider the number of supply points and when we refer in the following to the size of an energy provider we refer to the definition above. A comparison with the study from Müller et al. [5] shows that we had more smaller energy providers than they had considering the number of supply points as well as the number of employees.

To test similarity of the data for 2016 and 2018, we conducted a two-one-sided t-test (TOST) [14] for the energy provider's size and since for $\epsilon = 0.5$ the



(a) 2016 (b) 2018

Figure 2: Perceived Security

p-value of 0.027 was within the 95% confidence interval, we assume that the participating energy providers are similarly distributed within both surveys.

4 Results

Due to space limitations, we can only present an analysis of selected items of the questionnaire. We asked the participants about their perceived protection of systems and data in their company (cf. Fig. 2.) A Spearman's rank correlation test showed no correlation between size (cf. A2 in questionnaire) and perceived security (cf. B8 in questionnaire), but an independent-samples t-test (t(140)=2.5982, p-value = 0.01) suggests that the perceived security increased significantly[10] from 2016 to 2018.

4.1 ISMS Introduction

Tab. 1 shows that as expected, energy providers were quite active from 2016 to 2018 in implementing an ISMS (cf. C1 in questionnaire). While in 2016 75% of the energy providers only had at most 3 phases finished, in 2018 roughly

[10] mean in 2016: 2.41, in 2018: 2.06 with very good as 1 and sufficient as 4

half of them had 15 or more phases finished. This is also reflected in the mean 2.22 vs. 14.04 with a similar standard deviation (sd) and interquartile range (IQR). The status of the different ISMS implementation phases is shown in Fig. 3 which shows that besides the incident-management support most implementation phases a finished by are large majority.

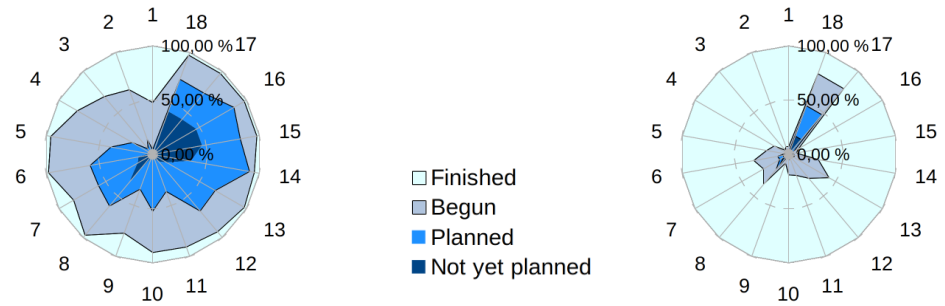
A Spearman's rank correlation test between the perceived security (cf. B8 in the questionnaire) and the number of finished ISMS phases (cf. C4 in the questionnaire) suggests also a significant small correlation (ρ-value: -0.27, p-value = 0.006). However, since the correlation was not significant when only considering the data from 2016 or 2018, we assume that this effect is merely the result of an increase in perceived security and increase of finished ISMS phases from 2016 to 2018.

Table 1: Distribution of finished ISMS implementation phases

Year	mean	sd	IQR	0%	25%	50%	75%	100%	n	NA
2016	2.22	3.12	3	0	0	1	3	17	46	15
2018	14.04	4.07	4	3	13	15	17	18	57	24

4.2 Motivation and Benefits from the ISMS

Figure 4a shows the energy providers' expectation of the effects of the ISMS's implementation along with the perceived benefits in 2018 (B) and the expected benefits in the future also in 2018 (E). It is visible, that for each of the reasons the energy providers expectations were outperformed. Figure 4b shows the result of the question why the energy providers had introduced an ISMS (in 2018). In both years legal requirements dominate the energy providers' motivation. We also asked in 2018 if the ISMS could improve the information



(a) 2016 (b) 2018

1 Target Setting and Scoping
2 ISMS Policy Development
3 Overview of the existing security architecture
4 Performing Risk Analyses
5 Elaboration of Catalogue of Security Measures
6 Design of the New Security Architecture
7 Description of Quality and Risk Manag. Interf.
8 Development of a Migration Process
9 Elaboration of the Required Documentation
10 Structure of the Security Organisation
11 Implementation of Management Processes
12 Formulation of Security Architecture (Rules)
13 Measures of Sensitization and Training
14 Implementation of Security Measures
15 Final Project Scope Analysis
16 Preparation for Certification Auditing
17 Execution of Business and Organisational Audits
18 Incident-Management Support

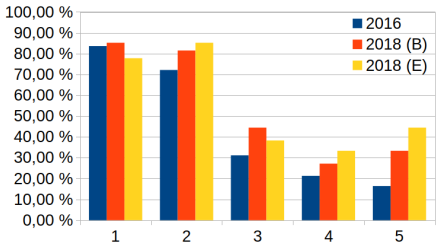
Figure 3: Status of each ISMS implementation phase

security and 93% confirmed that.

4.3 Effort of the ISMS Implementation

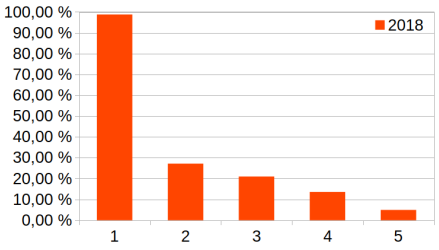
Table 2 shows the costs of the initial implementation of the ISMS (Tab. 2a) and of running the ISMS (Tab. 2b) divided into internal and external costs. Non surprisingly with increasing size, the costs also increase with the exception that the medium sized energy provider seem to have higher costs than large energy provider. The reason is that one medium provider reported very high costs (cf. maximum (100%) column). However, the Spearman's rank correlation test still suggests that there are moderate correlations between size and costs (for all 4 cost types, we found ρ-values between 0.44 and 0.53 with p-values below 10^{-3}). In 2016 (2018) 87% (96%) of the energy providers

1 Fulfilling Legal Requirements
2 Improving Information Security
3 Better Representation of IT Processes (Transparency)
4 Better External Representation of IT Processes
5 (Re-)Structuring of Relevant Business Processes



(a) Top 5 Reasons 2016 + Benefits and Future Expectations 2018

1 Legal Requirements
2 Business Processes are Depending on IT
3 Increased Threats
4 Public Discussion on IT-Security
5 Outsourcing of Services



(b) Top 5 Reasons 2018

Figure 4: Motivation, Benefits and Expectations to Implement an ISMS

reported that external consultants we supporting the implementation of the ISMS. However, only 55% reported that they will get external support for running and improving the ISMS.

4.4 Duration

Figure 5 shows the planned duration and the real duration of the ISMS implementation in months. While the duration seems to increase with the size, for medium sized energy providers (size 2), the range seems to be extremely large. We found a medium sized correlation between planned and real duration (0.61 with p-value $< 10^{-8}$), but Spearman's rank correlation suggests only a small correlation between planned (real) duration and energy provider size with ρ-value 0.27 (0.23) and p-value 0.02 (0.04). Overall, the mean real duration (20.7 months) is roughly 20% larger than the mean planned duration (17.0 months)

Table 2: Costs of the ISMS implementation (2018)

(a) Initial Costs

	Size	mean	sd	IQR	0%	25%	50%	75%	100%	n
Internal	S	56823	75707	50000	3000	10000	30000	60000	300000	17
	M	180275	504143	35525	10080	30000	50000	65525	2000000	15
	L	110000	64142	90000	30000	60000	80000	150000	250000	15
	XL	313500	543240	146500	30000	87500	150000	234000	2000000	12
External	S	54058	50380	60000	4000	20000	40000	80000	200000	17
	M	115891	245959	50000	20000	30000	45000	80000	1000000	15
	L	102058	53620	65000	25000	60000	100000	125000	220000	17
	XL	132769	97367	90000	25000	60000	110000	150000	350000	13

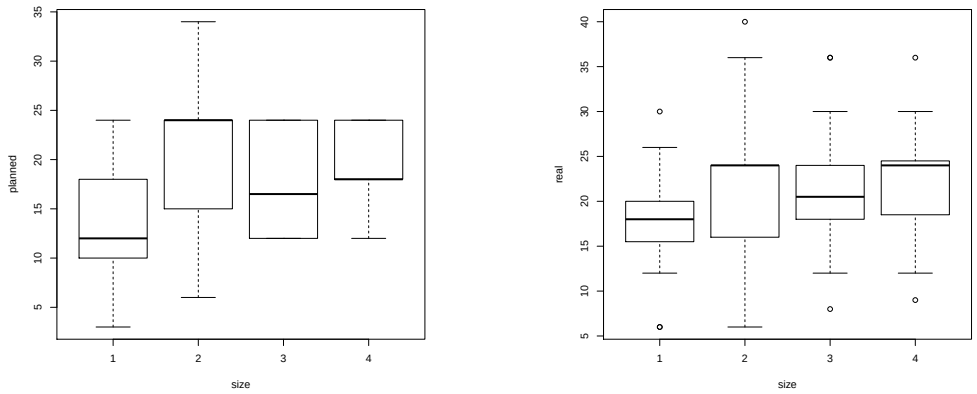
(b) Running Costs

	Size	mean	sd	IQR	0%	25%	50%	75%	100%	n
Internal	S	18529	16789	25000	1000	5000	10000	30000	50000	17
	M	72621	201748	17500	4320	10000	20000	27500	800000	15
	L	33000	23207	25000	10000	20000	25000	45000	100000	15
	XL	101538	126678	70000	10000	30000	80000	100000	500000	13
External	S	10000	12303	7625	1000	2375	6500	10000	50000	16
	M	28125	47314	10000	5000	10000	15000	20000	200000	16
	L	21866	13968	12500	5000	15000	20000	27500	50000	15
	XL	42285	48445	32500	5000	15000	35000	47500	200000	14

S: small; M: medium; L: large; XL: very large

5 Discussion

Results show that the perceived security was increased while in the same time almost all energy provider finished the implementation of their ISMS. This is in line with Müller et al. [5] who reported that 88% of the respondents had already implemented an ISMS. The latter is no surprise, given that the energy providers were legally obliged to do so, although we are aware that some of the small energy providers spent quite some effort to demonstrate that they do not fulfill the definition of a critical infrastructure, and thus do not need to implement and ISMS and get a corresponding certification. This matched the observation that most energy providers' main reason to implement an



(a) Planned Duration (b) Real Duration

Figure 5: Duration to Implement an ISMS (2018)

ISMS were legal requirements, which was also found by Müller et al. [5] (95%). Interestingly, while many were also expecting an increased information security, most of the energy providers had not started to implement an ISMS until they were required by law. On the other hand, more energy providers reported that their information security could benefit from the ISMS than the percentage of providers who expected that before. Again, this is in line with Müller et al. [5] who reported that for 95% the ISMS was beneficial for the energy provider.

Non surprisingly, larger energy providers reported higher costs for implementing and running the ISMS. It would have been interesting to compare that costs not only to the size but to the turnover. However, since many of the energy providers publish their balance sheets, we did not ask for it to ensure their anonymity. Müller et al. [5] reported a lower number of energy providers who got support from external consultants for implementing and

running the ISMS than we found, but since they had less smaller energy providers in their sample that most likely explains the difference.

5.1 Limitations

Although we checked for several reliability and validity issues, certain limitations might impact our results. First, the sample size can be considered relatively small for a quantitative study. However, since we checked all results for significance, we argue that our results are still valid, even though, we might have missed results with only a smaller effect size. Furthermore, it is difficult to gather data from energy providers since we could offer them no further incentive than the result of the study and their number is limited (roughly 900).

Our results face also possible self-selection biases since especially in 2018 energy providers who did not manage to implement a reasonable status of their ISMS might not have participated in the study. Additionally, since we decided to do the study anonymously, we could not link the participants from 2016 and 2018. This was an intentional decision, as we noticed that most energy providers were tense. Mainly because in 2016 the energy providers were not sure, what exactly they were required to do and in 2018 because they just had certified their ISMS or were still in the process of doing so.

6 Conclusion and Future Work

Our study suggests that information security of the energy providers benefits from the legislator's decision to require them to implement an ISMS (along

with other requirements). Most of the energy providers had not started and only implemented it when they were obliged to do so. The regulation also ensures fairness since all energy providers of a certain size are considered to be critical infrastructure, and thus need to implement it.

If would be interesting in future work to investigate in more detail how the energy providers are coping with new technology such as smart grids and virtual power plants. Furthermore, after the initial implementation, it will be interesting to observe how the energy providers cope with running the ISMS in a useful way.

7 Achknowledgements

This research was supported by the German Federal Ministry of Education within the project SIDATE (grant numbers: 16KIS0239K, 16KIS0240) and by the European Union within the project CyberSec4Europe (grant number 830929). We especially thank Benedikt Ley, Julian Dax and Frank Terhaag for their contribution to conduct the survey.

References

[1] D. Kipker, The EU NIS directive compared to the IT security act - Germany is well positioned for the new European cybersecurity space.

[2] B. (BSI), KRITIS-Sektorstudie Energie, Tech. rep. (02 2015).

[3] W. Hurst, M. Merabti, P. Fergus, A survey of critical infrastructure security, in: International Conference on Critical Infrastructure Protection, Springer, 2014, pp. 127–138.

[4] T. Rehbohm, K. Sandkuhl, T. Kemmerich, On challenges of cyber and information security management in federal structures-the example of german public administration, in: 2019 Joint International Conference on Perspectives in Business Informatics Research Workshops and Doctoral Consortium, BIR-WS 2019; Centre for New Information Technologies (CNTI), University of Economics in Katowice, Poland, 23-25 September 2019, Vol. 2443, CEUR-WS, 2019, pp. 1–13.

[5] J. Müller, A. Sänn, M. M. Wendt, R. A. Albrecht, P. Langendörfer, Informationssicherheits-management-systeme (isms) bei energieversorgern 2018.

[6] J. Dax, D. Hamburg, S. Pape, V. Pipek, K. Rannenberg, C. Schmitz, A. Sekulla, F. Terhaag, Sichere informationsnetze bei kleinen und mittleren energieversorgern (sidate), in: S. Rudel, U. Lechner (Eds.), State of the Art: IT-Sicherheit für Kritische Infrastrukturen, Universität der Bundeswehr, Neubiberg, 2018, Ch. Sichere Informationsnetze bei kleinen und mittleren Energieversorgern (SIDATE), p. 29.

[7] J. Dax, B. Ley, S. Pape, C. Schmitz, V. Pipek, K. Rannenberg, Elicitation of requirements for an inter-organizational platform to support security management decisions, in: 10th International Symposium on Human

Aspects of Information Security & Assurance, HAISA 2016 ,Frankfurt, Germany, July 19-21, 2016, Proceedings., 2016.

[8] C. Schmitz, A. Sekula, S. Pape, V. Pipek, K. Rannenberg, Easing the burden of security self-assessments, in: 12th International Symposium on Human Aspects of Information Security & Assurance, HAISA 2018 ,Dundee, Scotland, August 29-31, 2018, Proceedings., 2018.

[9] S. Pape, V. Pipek, K. Rannenberg, C. Schmitz, A. Sekulla, F. Terhaag, Stand zur IT-Sicherheit deutscher Stromnetzbetreiber : technischer Bericht (December 2018).

[10] J. Dax, B. Ley, S. Pape, V. Pipek, K. Rannenberg, C. Schmitz, A. Sekulla, Stand der it-sicherheit bei deutschen stromnetzbetreibern, in: S. Rudel, U. Lechner (Eds.), State of the Art: IT-Sicherheit für Kritische Infrastrukturen, Universität der Bundeswehr, Neubiberg, 2018, Ch. Stand der IT-Sicherheit bei deutschen Stromnetzbetreibern, pp. 69–74.

[11] J. Dax, A. Ivan, B. Ley, S. Pape, V. Pipek, K. Rannenberg, C. Schmitz, A. Sekulla, IT Security Status of German Energy Providers, also available via URN: urn:nbn:de:hbz:467-12141, URL: https://dokumentix.ub.uni-siegen.de/opus/volltexte/2017/1214/ (Sep. 2017). arXiv:arXiv:1709.01254.

[12] Bundesnetzagentur, Listen der Netzbetreiber und Versorgungsunternehmen (December 2019).

[13] Bundesnetzagentur, About us (September 2013).

[14] D. J. Schuirmann, A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability, Journal of pharmacokinetics and biopharmaceutics 15 (6) (1987) 657–680.

A Questionnaires

Question codes: ❶: question only appears in the 1st questionnaire; ❷: question only appears in the 2nd questionnaire.
Answer codes: ❖: multiple selection possible; ★: answers: "yes", "no" and "I don't know; ⊙: additional answer: "other"; ◘: additional answer: "I don't know".

A: General Company Information

A1 How many employees does your organisation have?
- Less than 50
- 51-100
- 101-250
- 251-500
- 501-1000
- More than 1000

A2 How many meter points are in your network?◘
- 0 - 1,000
- 1,001 - 15,000
- 15,001 - 30,000
- 30,001 - 100,000
- 100,001 - 500,000
- > 500,000

❶**A3** Which unbundling model is implemented in your company?◘⊙
- Small grid
- Major grid
- Lease
- No own grid

B: Organisational Aspects

B1 To which department are you assigned in the company?[11]

B2 What is your role in the company?

❶**B3** For how many employees in your company is IT security part of their daily business?

❷**B4** Who in your company is responsible for the operation of the ISMS?

❶**B5** Are there independent service providers in the field of IT security in your company? ★

❶**B6** Who takes on the role of IT security officer in your organization?◘⊙
- I myself
- Other employee
- External service provider
- There is no

❶**B7** To which department is the IT security officer assigned? [11]

B8 In your view, how well protected are the systems and data in your company?◘
- Very good
- Good
- Satisfying
- Sufficient

C: ISMS

❶**C1** The introduction of an ISMS is/has ... ◘
- not planned yet
- planned
- already started
- already completed

❶**C2** When should the work on the introduction of an ISMS begin or when did it start?[12]

C3 When was the work on introducing an ISMS completed?[13]

C4 What is the current status of the respective ISMS implementation phases?[14]

❶**C5** By when should the work on introducing an ISMS be completed? [15]

C6 When was the work on introducing an ISMS completed? [16]

C7 How long did you expect the introduction of an ISMS to take at the beginning of the implementation?

C8 How long did it actually take to implement your ISMS?

C9* Have external service providers been or will be consulted when introducing an ISMS?

C10 What were the main reasons for you to introduce an ISMS?❖
- Legal requirements (IT security cata-

[11] Answer options were: Management IT, Power system management Administration & organization, Legal department, Public relations, Other

[12] Answer options: half years from 2nd 2016 to 2nd 2018, Later and I don't know

[13] Answer options: half years from 2nd 2013 to 2nd 2018, Earlier and I don't know

[14] It has been asked for the current status (not yet planned, planned, begun, or finished) for the implementation phases described in Fig. 3

[15] Answer options: half years from 2nd 2016 to 2nd 2019, Later and I don't know

[16] Answer options: half years from 2nd 2013 to 2nd 2016, Earlier and I don't know

logue, IT security law)
- Increased threat level
- Strong dependence of business operations on IT
- Outsourcing of services to external service providers
- Public discussion on IT security

❷**C11** In which areas have you already been able to benefit from the introduction of the ISMS?❖◘⊙[17]

C12 What do you hope for or expect from the introduction of an ISMS?❖◘⊙[17]

D: ISMS Maintenance

❷**D1** In your opinion, could the security level of your company be improved by implementing the ISMS?★

❷**D2** How high were your initial costs for the introduction of the ISMS?★

❷**D3** Do you have continuous external support for the operation of the ISMS?★

❷**D4** What annual costs do you expect for the operation of your ISMS?★
- Internal costs
- External costs

❷**D5** In which areas of ISMS operation are the greatest challenges for your company?⊙
- Technical adjustments
- Adaptation of procedures/processes
- Lack of personnel
- Missing hardware
- Process monitoring
- Documentation
- Cooperation with external
- Risk Management
- Implementation of continuous safety improvement

❷**D6** Work together with other network operators in the field of ISMS operation or exchange information with from other network operators?★
- Regular cooperation with other network operators
- Regular exchange with other network operators
- Occasional exchange with other network operators
- Little or no exchange with other network operators

❷**D7** In your opinion, could cooperation with other network operators contribute to the operation of your ISMS or security level?★

E: Office IT

E1 Are there IT security guidelines for the office IT in your company?★

E2 Are the IT security guidelines updated and, if necessary, adjusted regularly?★

F: ICS[18]: network structure

F1 Does your energy control system enables only energy network supervision, or does it also enable to execute switching operations?◘
- Supervision only
- Supervision and control

F2 How is the IT network of your energy control system separated from other networks (e.g. IT department, Internet, maintenance companies)?◘
- Logical Separation
- Physical Separation
- No Separation

F3 Is the network of your energy control system divided in different security domains (e.g. through different VLANs)?★

❶**F4** Which network technologies do you use in your energy control system network?❖◘
- Cable connect.
- Wireless connect.

❶**F5** Which communication standards are used in the network of your energy control system?

❶**F6** What wireless network technologies do you use?

❶**F7** Which communication standards are used in your control system

[17] Answer options were: Improvement of information security in the company, (Re)structuring of the relevant business processes,Legal compliance, Better representation of IT processes,Better external presentation of the IT security processes.

[18] Industrial Control System

network?❖□⊙
- IP communication
- Serial communication

❶**F8** From which producers do you acquire the network administration systems and devices?

F9 Which types of remote access were established for your energy control system?❖□⊙
- External Access for maintenance and configuration of the control system
- Employee Access (e.g. for standby or fault clearance service)

F10 How are remote access procedures via external service providers regulated?⊙
- External service providers can have access to the system and undertake changes only after receiving authorization, but WITHOUT additional surveillance
- External service providers can have access to the system and undertake changes only after receiving authorization and only under surveillance
- External service providers can have access to the system and undertake changes independently

G: ICS: Processes and Organisation

G1 Are you/the responsible employees regularly informed about potential hard-/software vulnerabilities?★

G2 How often are the devices and software within your energy control system updated/renewed?[19]

G3 Is there an updated inventory list in which all the software items are documented (e.g. with version numbers, corresponding accounts and IP addresses)?★

G4 Are there documented IT security guidelines for the energy control system in your company?★

G5 Under which security-relevant standards are your IT systems and processes for network administration elaborated?□⊙
- ISO/IEC 27001
- None
- BSI Grundschutz

G6 Do you perform IT risk analyses for the processes and IT systems for network administration?★

G7 How often do you perform such risk analyses?[20]

G8 Do you perform security audits, vulnerability scans, or penetration tests for the administration systems of the network management technology?□
- Yes; by external service providers
- Yes; by own employees
- Yes; by both external providers and employees
- No

G9 How often do you perform such vulnerability scans or penetration tests?[20]

G10 Do you have an emergency plan for security incidents of network administration?★

G11 Are security-relevant incidents (e.g. portscans, failed login attempts, unauthorised processes) recorded and evaluated?□
- Yes, only logging
- Yes, logging and evaluation
- No, neither

G12 Which information do you evaluate to identify attacks on the IT systems for network control?❖□⊙
- Firewall logs
- Failed logins
- System logs
- Honeypot logs

G13 Do you use metrics to assess vulnerabilities (e.g. CVSS)?★

G14 Is IT security defined as a requirement for acquiring new hard- and software?★

[19]It has been asked for the update frequency (regularly, for known vulnerabilities, not yet, or I don't know) of: network equipment (e.g. routers, switches), workstation computer/terminal, server, and network control/telecontrol technology

[20]Answer options were: More than once a year,yearly, every two years, more rarely, I don't know

Appendix C

Privacy Enhancing Technologies

C.1 Examining Technology Use Factors of Privacy-Enhancing Technologies: The Role of Perceived Anonymity and Trust

David Harborth and Sebastian Pape. Examining technology use factors of privacy-enhancing technologies: The role of perceived anonymity and trust. In *24th Americas Conference on Information Systems, AMCIS 2018, New Orleans, LA, USA, August 16-18, 2018*. Association for Information Systems, 2018. URL https://aisel.aisnet.org/amcis2018/Security/Presentations/15

Examining Technology Use Factors of Privacy-Enhancing Technologies: The Role of Perceived Anonymity and Trust

Completed Research

David Harborth
Chair of Mobile Business & Multilateral Security
Goethe University Frankfurt
david.harborth@m-chair.de

Sebastian Pape
Chair of Mobile Business & Multilateral Security
Goethe University Frankfurt
sebastian.pape@m-chair.de

Abstract

Today's environment of data-driven business models relies heavily on collecting as much personal data as possible. This is one of the main causes for the importance of privacy-enhancing technologies (PETs) to protect internet users' privacy. Still, PETs are rather a niche product used by relatively few users on the internet. We undertake a first step towards understanding the use behavior of such technologies. For that purpose, we conducted an online survey with 141 users of the anonymity service "JonDonym". We use the technology acceptance model as a theoretical starting point and extend it with the constructs perceived anonymity and trust in the service. Our model explains almost half of the variance of the behavioral intention to use JonDonym and the actual use behavior. In addition, the results indicate that both added variables are highly relevant factors in the path model.

Keywords

Privacy-enhancing technologies (PETs), technology use, technology acceptance, perceived anonymity, trust, privacy, structural equation model.

Introduction

Perry Barlow (Ball 2012) states: “The internet is the most liberating tool for humanity ever invented, and also the best for surveillance. It's not one or the other. It's both.” One of the reasons for surveilling users is a rising economic interest in the internet (Bédard 2016). However, users who have privacy concerns and feel a strong need to protect their privacy are not helpless, they can make use of privacy-enhancing technologies (PETs). PETs allow users to improve their privacy by eliminating or minimizing personal data disclosure to prevent unnecessary or unwanted processing of personal data (van Blarkom et al. 2003). PETs have a property that is not characteristic for many other technology types. They usually serve not only the primary goals of the users, but also their secondary goals (Cranor and Garfinkel 2008). It is important to understand that in many cases PET users make use of the PET while they pursue another goal like browsing the internet or using instant messengers. These aims become more indistinct if the PET is integrated in the regular service (e.g. anonymous credentials (Benenson et al. 2015)). In contrast to PETs integrated in services, standalone PETs (e.g. overlay networks like Tor (The Tor Project 2018)) are not integrated into a specific service and can be used for several purposes.

In this paper, we investigate how the users’ main goal (privacy respectively anonymity) and their trust in the service influence the intention to use the PET. In order to focus on the PET itself and not to interfere with possible other goals, we choose a standalone PET as object for investigation. This allows us to focus on the usefulness of the PET with regard to privacy protection and avoids confounders due to other goals of the user. Therefore, we conducted a survey of the users of the anonymity service JonDonym. JonDonym is a proxy client and will forward the traffic of the users’ internet applications encrypted to the mix cascades to hide their IP addresses (JonDos Gmbh 2018).

To determine the use factors of this PET, we focused on perceived anonymity and trust: Since most users do not base their decisions on any kind of formal (technical or mathematical) anonymity measurement, we decided to measure the perceived anonymity. The resulting research question is:

RQ1: Does perceived anonymity influence the behavioral intention to use a PET?

However, perceived anonymity is a subjective perception of each user. Since we assume, that most users will not dig into mathematical proofs of the assured anonymity or challenge the implementation of the service provider, we conclude that it is important to also consider the trust in the service provider and the service itself:

RQ2: Does trust in the PET influences the behavioral intention to use it?

We further refine the two research questions and in particular the connection between perceived anonymity, perceived usefulness and trust in the service (JonDonym) in section 3. This allows us to integrate them into a technology acceptance model (TAM) which we then use to answer the research questions.

The remainder of the paper is structured as follows: Section 2 briefly introduces the JonDonym anonymization service and lists related work on PETs and technology acceptance. In section 3, we present the research hypotheses and describe the questionnaire and the data collection process. We assess the quality of our empirical results with regard to reliability and validity in section 4. In section 5, we discuss the implications of the results, elaborate on limitations of our work and conclude the paper with suggestions for future work.

Background and Related Work

Privacy-Enhancing Technologies (PETs) is an umbrella term for different privacy protecting technologies. Borking and Raab define PETs as a "coherent system of ICT measures that protects privacy [...] by eliminating or reducing personal data or by preventing unnecessary and/or undesired processing of personal data; all without losing the functionality of the data system" (Borking and Raab 2001, p. 1).

In this paper, we investigate the role of perceived anonymity and trust in the context of a technology acceptance model for the case of the anonymity service JonDonym (JonDos Gmbh 2018). Comparable to Tor (The Tor Project 2018), JonDonym is an anonymity service and a PET. However, unlike Tor, it is a proxy system based on mix cascades. It is available for free with several limitations, like a restricted maximum download speed. In addition, there are different premium rates without these limitations that differ with regard to duration and included data volume. Thus, JonDonym offers several different tariffs and is not based on donations like Tor. The actual number of users is not predictable since the service does not keep track of this. JonDonym is also the focus of an earlier user study on user characteristics of privacy services (Spiekermann 2005). However, the focus of the study is rather descriptive and does not focus on users' beliefs and concerns.

Previous non-technical work on PETs considers mainly usability studies and does not primarily focus on privacy concerns and related trust and risk beliefs of PET users. For example, Lee et al. (2017) assess the usability of the Tor Launcher and propose recommendations to overcome the found usability issues. Comparable studies to ours are the ones by Benenson et al. (2014, 2015) and Krontiris et al. (2015), who investigate acceptance factors for an anonymous credentials service. However, in their case the anonymous credential service is integrated into an evaluation system. Thus, the users of their anonymous credential service had a clearly defined primary task (evaluation of the course system) and a secondary task (ensure privacy protection). Benenson et al. (2014) focused on the measurement of the perceived usefulness of the anonymous credential system (the secondary goal), but state that considering the perceived usefulness for the primary goals as well, may change the relationship between the variables in their model. In contrast to their study, we examine a standalone PET, and thus can focus on privacy protection as the primary goal of the users with respect to the PET.

Methodology

We base our research on the well-known technology acceptance model (TAM) by Davis (1985, 1989). For analyzing the cause-effect relationships between the latent variables, we use structural equation modelling (SEM). There are two main approaches for SEM, namely covariance-based SEM (CB-SEM) and partial least squares SEM (PLS-SEM). Since our research goal is to predict the target construct actual use behavior of JonDonym, we use PLS-SEM for our analysis (Hair et al. 2011). In the following subsections, we discuss the research model and hypotheses based on the extended TAM, the questionnaire and the data collection process. The demographic questions were not mandatory to fill out. This was done on purpose since we assumed that most of the participants are highly sensitive with respect to their personal data. Therefore, we resign from a discussion of the demographics in our research context. This decision is backed up by Singh and Hill, who found no statistically significant differences across gender, income groups, educational levels, or political affiliation in the desire to protect one's privacy (Singh and Hill 2003).

Research Model and Hypotheses

PETs are structurally different than formerly investigated technologies in the job context or hedonic information systems. In general, it is obvious to users what a certain technology does. For example, if users employ a spreadsheet program in their job environment, they will see the immediate result of their action when the program provides them a calculation. The same holds for hedonic technologies which provide an immediate feedback to the user during the interaction. However, this interaction and feedback structure is different with PETs. The main impact a user can achieve by using JonDonym is anonymity. However, most PETs are designed to not harm the user experience. Besides some negative side effects such as a loss of speed during browsing the internet or an increasing occurrence of captchas (Chirgwin 2016), the user may not be able to detect the PET at all. The direct effects of the increased anonymity in general go undetected since they consist of long term consequences, e.g. different advertisements, unless the user visits special websites with anonymity tests or showing the internet address of the request.

Therefore, perceptions about the achieved impact of using the technology should be specifically incorporated in any model dealing with drivers of use behavior. This matches the observation that most users do not base their decisions on any kind of formal (technical or mathematical) anonymity measurement. Thus, we adapted a formerly validated construct named "perceived anonymity" to the case of JonDonym (Benenson et al. 2015). The construct mainly asks for the perceptions of users about their level of anonymity achieved by the use of the PET. Due to the natural importance of anonymity for a PET, we argue that these perceptions will have an important effect on the trust in the technology. Thus, the more users think that the PET will create anonymity during their online activities, the more they will trust the PET (H1a). Creating anonymity for its users is the main use of a PET. Thus, we hypothesize that the perceived anonymity has a positive effect on the perceived usefulness of the PET to protect the user's privacy (H1b).

H1a: Perceived anonymity achieved by using JonDonym has a positive effect on trust in JonDonym.

H1b: Perceived anonymity achieved by using JonDonym has a positive effect on the perceived usefulness of JonDonym to protect the user's privacy.

Trust is a diverse concept integrated in several models in the IS domain. It is shown that different trust relationships exist in the context of technology adoption of information systems (Söllner et al. 2016). Trust can refer to the technology (in our case JonDonym) itself as well as to the service provider (in our case JonDos). However, JonDonym is the company's main product. Therefore, we argue that it is rather difficult for users to distinguish which label refers to the technology itself and which refers to the company. Thus, we decided to ask for trust in the service (JonDonym), assuming that the difference to ask for trust in the company is negligible. The items for measuring trust and the effects of trust on other variables of the technology acceptance model are adapted from Pavlou (2003). Thus, we hypothesize that trust influences behavioral intention, perceived usefulness and perceived ease of use positively.

H2a: Trust in JonDonym has a positive effect on the behavioral intention to use the technology.

H2b: Trust in JonDonym has a positive effect on the perceived usefulness of JonDonym to protect the user's privacy.

H2c: Trust in JonDonym has a positive effect on the perceived ease of use of JonDonym.

The theoretical underlying of hypotheses H3, H4a, H4b and H5 can be adapted from the original work on TAM by Davis (1985, 1989) since PETs are not different to other technologies with regard to relationships of perceived usefulness, perceived ease, behavioral intention to use and actual use behavior. However, perceived usefulness refers explicitly to privacy protection as it is the sole purpose of the technology. Thus, we hypothesize:

H3: The perceived usefulness of protecting the user's privacy has a positive effect on the behavioral intention to use the technology.

H4a: Perceived ease of use has a positive effect on the behavioral intention to use the technology.

H4b: Perceived ease of use has a positive effect on the perceived usefulness of JonDonym to protect the user's privacy.

H5: The behavioral intention to use JonDonym has a positive effect on the actual use behavior.

Questionnaire Composition and Data Collection Procedure

The questionnaire constructs are adapted from different sources. The constructs Perceived ease of use (PEOU) and perceived usefulness are adapted from Venkatesh and Davis (2000), behavioral intention (BI) is adapted from Venkatesh et al. (2012), trust in the PET service is adapted from Pavlou (2003) and perceived anonymity is adapted from Benenson et al. (2015). The actual use behavior is measured with a ten-item frequency scale (Rosen et al. 2013). We conducted the study with German and English speaking JonDonym users. Thus, we administered two questionnaires. All items for the German questionnaire had to be translated into German since all of the constructs are adapted from English literature.

To ensure content validity of the translation, we followed a rigorous translation process. First, we translated the English questionnaire into German with the help of a certified translator (translators are standardized following the DIN EN 15038 norm). The German version was then given to a second independent certified translator who retranslated the questionnaire to English. This step was done to ensure the equivalence of the translation. Third, a group of five academic colleagues checked the two English versions with regard to this equivalence. All items were found to be equivalent. The items can be found in Table 1.

Since we investigate the drivers of the use behavior of JonDonym, we collected data from actual users of the PET. We installed the surveys on a university server and managed it with the survey software LimeSurvey (version 2.63.1) (Schmitz 2015). The links to the English and German version were distributed with the beta version of the JonDonym browser and published on the official JonDonym homepage. This made it possible to address the actual users of the PET in the most efficient manner. In sum, 416 participants started the questionnaire (173 for the English version and 243 for the German version). Of those 416 approached participants, 141 (53 for the English version and 88 for the German version) remained after deleting unfinished sets and all participants who answered a test question in the middle of the survey incorrectly.

Results

We tested the model using SmartPLS version 3.2.7 (Ringle et al. 2015). Before looking at the result of the structural model and discussing its implications, we discuss the measurement model, and check for the reliability and validity of our results. This is a precondition of being able to interpret the results of the structural model. Furthermore, it is recommended to report the computational settings. For the PLS algorithm, we choose the path weighting scheme with a maximum of 300 iterations and a stop criterion of 10^{-7}. For the bootstrapping procedure, we use 5000 bootstrap subsamples and no sign changes as the method for handling sign changes during the iterations of the bootstrapping procedure. In addition, it is relevant to mention that we met the suggested minimum sample size with 141 datasets considering the threshold of ten times the number of structural paths headed towards a latent construct in the model (Hair et al. 2011).

Measurement Model Assessment

As the model is measured solely reflectively, we need to evaluate the internal consistency reliability, convergent validity and discriminant validity to assess the measurement model properly (Hair et al. 2011). Internal consistency reliability (ICR) measurements indicate how well certain indicators of a construct measure the same latent phenomenon. Two standard approaches for assessing ICR are Cronbach's ∝ and the composite reliability. The values of both measures should be between 0.7 and 0.95 for research that builds upon accepted models. Values of Cronbach's α are seen as a lower bound and values of the composite reliability as an upper bound of the assessment (Hair et al. 2017). Table 1 includes the ICR of the variables in the last two rows. It can be seen that all values for Cronbach's α and the composite reliability are above the lower threshold of 0.7 and no value is above 0.95. In sum, ICR is established for our variables.

Constructs	BI	PEOU	PA	Trust	PU
BI1. I intend to continue using JonDonym in the future.	**0.913**	0.432	0.546	0.622	0.541
BI2. I will always try to use JonDonym in my daily life.	**0.806**	0.328	0.331	0.362	0.313
BI3. I plan to continue to use JonDonym frequently.	**0.941**	0.393	0.466	0.582	0.458
PEUO1. My interaction with JonDonym is clear and understandable.	0.369	**0.862**	0.224	0.372	0.327
PEUO2. Interacting with JonDonym does not require a lot of my mental effort.	0.349	**0.843**	0.130	0.224	0.227
PEUO3. I find JonDonym to be easy to use.	0.341	**0.920**	0.145	0.246	0.303
PEUO4. I find it easy to get JonDonym to do what I want it to do.	0.444	**0.893**	0.373	0.426	0.464
PA1. JonDonym is able to protect my anonymity in during my online activities.	0.398	0.151	**0.882**	0.482	0.584
PA2. With JonDonym I obtain a sense of anonymity in my online activities.	0.489	0.254	**0.874**	0.593	0.657
PA3. JonDonym can prevent threats to my anonymity when being online.	0.445	0.297	**0.869**	0.480	0.574
Trust1. JonDonym is trustworthy.	0.494	0.321	0.580	**0.909**	0.557
Trust2. JonDonym keeps promises and commitments.	0.568	0.365	0.531	**0.922**	0.505
Trust3. I trust JonDonym because they keep my best interests in mind.	0.576	0.350	0.526	**0.911**	0.491
PU1. Using JonDonym improves the performance of my privacy protection.	0.330	0.347	0.553	0.398	**0.885**
PU2. Using JonDonym increases my level of privacy.	0.468	0.334	0.669	0.578	**0.923**
PU3. Using JonDonym enhances the effectiveness of my privacy.	0.304	0.322	0.547	0.372	**0.855**
PU4. I find JonDonym to be useful in protecting my privacy.	0.592	0.377	0.653	0.590	**0.863**
Cronbach's ∝	0.865	0.904	0.847	0.902	0.906
Composite Reliability	0.918	0.932	0.907	0.939	0.933

Table 1. Loadings and Cross-Loadings of the Reflective Items and ICR measures

In a next step, we assess the convergent validity to determine the degree to which indicators of a certain reflective construct are explained by that construct. For that, we calculate the outer loadings of the indicators of the constructs (indicator reliability) and evaluate the average variance extracted (AVE) (Hair et al. 2017). Loadings above 0.7 imply that the indicators have much in common, which is desirable for reflective measurement models. Table 1 shows the outer loadings in bold on the diagonal. All loadings are higher than 0.7. Convergent validity for the construct is assessed by the AVE. AVE is equal to the sum of the squared loadings divided by the number of indicators. A threshold of 0.5 is acceptable, indicating that the construct explains at least half of the variance of the indicators. The first column of Table 2 presents the AVE of the constructs. All values are well above 0.5, demonstrating convergent validity.

The next step for assessing the measurement model is the evaluation of discriminant validity. It measures the degree of uniqueness of a construct compared to other constructs. Comparable to the convergent validity assessment, two approaches are used for investigated discriminant validity. The first approach, assessing cross-loadings, is dealing with single indicators. All outer loadings of a certain construct should be larger than its cross-loadings with other constructs (Hair et al. 2017). Table 1 illustrates the cross-loadings as off-diagonal elements. All cross-loadings are smaller than the outer loadings, fulfilling the first assessment approach of discriminant validity. The second approach is on the construct level and compares the square root of the constructs' AVE with the correlations with other constructs. The square root of the AVE of a single construct should be larger than the correlation with other constructs (Fornell-Larcker criterion). Table 2 contains the square root of the AVE as on-diagonal values. All values are larger than the correlations with other constructs, indicating discriminant validity.

Constructs (AVE)	BI	PA	PEOU	PU	Trust
BI (0.790)	0.889				
PA (0.765)	0.510	0.875			
PEOU (0.774)	0.435	0.268	0.880		
PU (0.778)	0.500	0.695	0.393	0.882	
Trust (0.836)	0.597	0.597	0.378	0.566	0.914

Table 2. Discriminant Validity with AVEs and Construct Correlations

The last step of the measurement model assessment is the check for common method bias (CMB). CMB can occur if data is gathered with a self-reported survey at one point in time in one questionnaire (Malhotra et al. 2006). Since this is the case in our research design, the need to test for CMB arises. An unrotated principal component factor analysis is performed with the software package STATA 14.0 to conduct the Harman's single-factor test to address the issue of CMB (Podsakoff et al. 2003). The assumptions of the test are that CMB is not an issue if there is no single factor that results from the factor analysis or that the first factor does not account for the majority of the total variance. The test shows that four factors have eigenvalues larger than 1 which account for 75.48% of the total variance. The first factor explains 45.35% of the total variance. Thus, no single factor emerged and the first factor does not explain the majority of the variance. Hence, we argue that CMB is not likely to be an issue in the data set.

Structural Model Assessment

We first test for possible collinearity problems before discussing the results of the structural model. Collinearity is present if two predictor variables are highly correlated with each other. This is important since collinearity can otherwise bias the results heavily. To address this issue, we assess the inner variance inflation factor (inner VIF). All VIF values above 5 indicate that collinearity between constructs is present (Hair et al. 2017). For our model, the highest VIF is 1.688. Thus, collinearity is apparently not an issue.

Figure 1 presents the results of the path estimations and the R^2-values of the target variables behavioral intention and actual use behavior. In addition, we provide the R^2-values for trust, perceived ease of use and perceived usefulness. R^2-values are weak with values around 0.25, moderate with 0.50 and substantial with 0.75 (Hair et al. 2011). Based on this classification, the R^2-values for behavioral intention and actual use are rather moderate in size. Thus, our model explains 42.9% of the variance in the behavioral intention to use the PET and 46.1% of the variance of the actual use behavior. This result is very good considering the parsimonious measurement model. In addition, the explained variance of perceived usefulness is 54.7%, indicating that the three variables, perceived anonymity, trust and perceived ease of use explain more than half of the variance of this construct.

Thus, we identified three major drivers of users' perceptions with regard to the usefulness of a privacy-enhancing technology. The strongest effect is exerted by the users' perceived anonymity provided by the service (H1b confirmed). This result is not surprising considering that providing anonymity is the main goal of a PET. In addition, perceived anonymity has a strong and statistically significant effect on trust (H1a confirmed). Thus, users' trust in the PET is mainly driven by their perceptions that the service can create anonymity.

As hypothesized in H2a - H2c, trust has a significant positive effect on the behavioral intention to use the PET, the perceived usefulness and the perceived ease of use. Therefore, trust emerges as a highly relevant concept when determining the drivers of users' use behavior of PETs. It has the strongest effect size (0.416) on behavioral intention. As discussed earlier, hypotheses H3 - H5 are adapted from the original work on TAM (Davis 1985, 1989) and can be confirmed for the case of PETs.

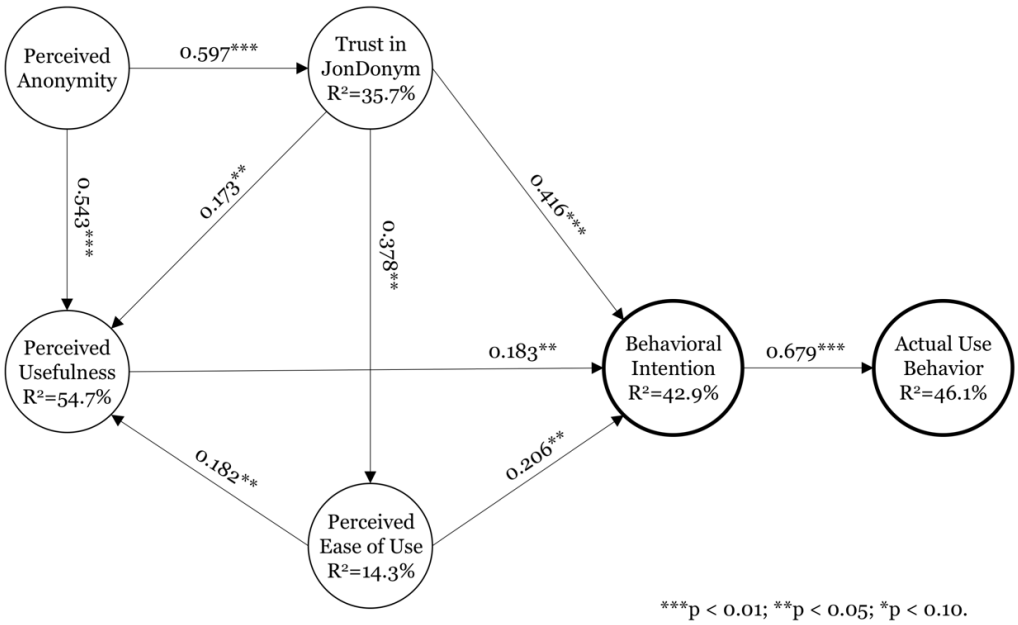


Figure 1. Path Estimates and Adjusted R^2-values of the Structural Model

Since the effects of perceived anonymity and trust on behavioral intention and the actual use behavior are partially indirect, we determine and analyze the total effects for these variables (cf. Table 3). It can be seen that all total effects are relatively large and highly statistically significant. Thus, perceived anonymity and trust strongly influence the target variables BI and USE.

Total effect	Effect size	P-value
PA → BI	0.431	0.000
PA → USE	0.289	0.000
Trust → BI	0.551	0.000
Trust → USE	0.370	0.000

Table 3. Total Effects for the Variables Perceived Anonymity and Trust

As a next, we assessed the predictive relevance of the two added variables for behavioral intention and actual use behavior. A simple measure for the relevance of perceived anonymity and trust is to delete both variables and run the model again. The results show that the R^2-value for behavioral intention decreases to 31.9% (= eleven percentage points less). Thus, without the two new variables the explained variance for behavioral intention decreases by roughly a quarter (25.64%). A more advanced measure for predictive relevance is the Q^2 measure. It indicates the out-of-sample predictive relevance of the structural model with regard to the endogenous latent variables based on a blindfolding procedure (Hair et al. 2017). We used an omission distance d=7. Recommended values for d are between five and ten. Furthermore, we report the Q^2 values of the cross-validated redundancy approach, since this approach is based on both the results of the measurement model as well as of the structural model. Detailed information about the calculation cannot be provided due to space limitations. For further information see Chin (1998). For our model, Q^2 is calculated for behavioral intention and use behavior. Values above 0 indicate that the model has the property of predictive relevance. Omitting both new variables leads to a decrease of Q^2 for behavioral intention from 0.304 to 0.223. R^2 as well as Q^2 did not change for actual use when deleting the new variables, since there is not direct relation from the constructs to the actual use construct and behavioral intention solely explains a large share of variance in use.

Discussion and Conclusion

Research on privacy-enhancing technologies mainly focused on the technical aspects of the technologies up to now. However, a successful implementation and adoption of PETs requires of profound understanding of the perceptions and behaviors of actual and possible users of the technologies. The IS domain has the proper methods and knowledge to tackle such questions. Thus, with this paper we investigated actual users of an existing PET as a first step to address this research problem. Our results indicate that the basic rationale of technology use models holds for privacy-enhancing technologies. However, the newly introduced variables perceived anonymity and trust strongly improved the explanatory of the structural model for the case of a PET and should be considered for comparable research problems in future work.

Although we checked for several reliability and validity issues, certain limitations might impact our results. First, the sample size of 141 participants is relatively small for a quantitative study. However, since we reached the suggested minimum sample size for the applied method, we argue that our results are still valid. In addition, it is very difficult to gather data of actual users of PETs since it is a comparable small population that we could survey. It is also relevant to mention that we did not offer any financial rewards for the participation. A second limitation concerns possible self-report biases (e.g. social desirability). We addressed this possible issue by gathering the data fully anonymized. Furthermore, demographic questions were not mandatory to fill out. Third, mixing results of the German and English questionnaire could be a source of errors. On the one hand, this procedure was necessary to achieve the minimum sample size. On the other hand, we followed a very thorough translation procedure to ensure the highest level of equivalence as possible. Thus, we argue that this limitation did not affect the results. Lastly, we did not control for the participants' actual or former use of different standalone PETs. This experience might have an impact on their assessments of JonDonym.

We found strong effects for the influence of the perceived anonymity on the behavioral intention to use a PET (RQ1). In contrast to the findings of Benenson et al. (2015), who found that trust in the PET has no

statistically significant impact on the intention to use the service, we also found a strong effect of trust in the PET on the behavioral intention to use it (RQ2). One reason for the difference might be that the trust in the service and the trust in the service provider were very likely equivalent in our use case. However, to adequately address the difference further research is needed. From a practical point of view, our results indicate that PET providers should aim to establish a trustworthy service with a high level of transparency in order to increase the perceived anonymity of users.

Future work can build on the proposed relationships and extensions of our model to investigate the acceptance and use of PETs in more detail. We could explain almost half of the variance in the target constructs behavioral intention and actual use behavior with a rather parsimonious model. Thus, the current model provides a good starting point to investigate other comparable PETs, like Tor or a VPN service. In addition, new privacy or technology-specific variables could be added to strengthen the understanding about usage of PETs. Based on our findings, future work could also investigate the found relationships with a qualitative research approach in more detail. In a next step, it would be interesting to investigate the perceptions of non-users about PETs and compare the findings to actual users. By that, it would be possible for developers and marketers to specifically address issue hindering a broader diffusion of PETs. This could be a real contribution for strengthening the personal right for privacy in times of ever-increasing personal data collection in the internet.

Acknowledgements

This research was partly funded by the German Federal Ministry of Education and Research (BMBF) with grant number: 16KIS0371. In addition, we thank Rolf Wendolsky (JonDos GmbH) for his help during the data collection process.

REFERENCES

Ball, J. 2012. "Hacktivists in the Frontline Battle for the Internet," *The Guardian*. (https://www.theguardian.com/technology/2012/apr/20/hacktivists-battle-internet, accessed February 26, 2018).

Bédard, M. 2016. "The Underestimated Economic Benefits of the Internet," in *Regulation Series, The Montreal Economic Institute*.

Benenson, Z., Girard, A., and Krontiris, I. 2015. "User Acceptance Factors for Anonymous Credentials: An Empirical Investigation," *14th Annual Workshop on the Economics of Information Security (WEIS)*, pp. 1–33.

Benenson, Z., Girard, A., Krontiris, I., Liagkou, V., Rannenberg, K., and Stamatiou, Y. C. 2014. "User Acceptance of Privacy-ABCs: An Exploratory Study," in *Human–Computer Interaction*, pp. 375–386.

van Blarkom, G. W., Borking, J. J., and Olk, J. G. E. 2003. *"PET". Handbook of Privacy and Privacy-Enhancing Technologies*.

Borking, J. J., and Raab, C. 2001. "Laws, PETs and Other Technologies for Privacy Protection," *Journal of Information, Law and Technology* (1), pp. 1–14.

Chin, W. W. 1998. "The Partial Least Squares Approach to Structural Equation Modeling," in *Modern Methods for Business Research*, G. A. Marcoulides (ed.), Mahwah, NJ: Lawrence Erlbaum, pp. 295–336.

Chirgwin, R. 2016. "CloudFlare Shows Tor Users the Way out of CAPTCHA Hell," *The Register*. (https://www.theregister.co.uk/2016/10/05/cloudflare_tor/, accessed February 23, 2018).

Cranor, L. F., and Garfinkel, S. 2008. *Security and Usability: Designing Secure Systems That People Can Use*, Farnham: O'Reilly.

Davis, F. D. 1985. "A Technology Acceptance Model for Empirically Testing New End-User Information Systems: Theory and Results," *Massachusetts Institute of Technology*.

Davis, F. D. 1989. "Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology," *MIS Quarterly* (13:3), pp. 319–340.

Hair, J., Hult, G. T. M., Ringle, C. M., and Sarstedt, M. 2017. *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)*, SAGE Publications.

Hair, J., Ringle, C. M., and Sarstedt, M. 2011. "PLS-SEM: Indeed a Silver Bullet," *The Journal of Marketing Theory and Practice* (19:2), pp. 139–152.

JonDos Gmbh. 2018. "Official Homepage of JonDonym." (https://www.anonym-surfen.de, accessed

January 16, 2018).
Krontiris, I., Benenson, Z., Girard, A., Sabouri, A., Rannenberg, K., and Schoo, P. 2015. "Privacy-ABCs as a Case for Studying the Adoption of PETs by Users and Service Providers," in *APF*, pp. 104–123.
Lee, L., Fifield, D., Malkin, N., Iyer, G., Egelman, S., and Wagner, D. 2017. "A Usability Evaluation of Tor Launcher," *Proceedings on Privacy Enhancing Technologies* (3), pp. 90–109.
Malhotra, N. K., Kim, S. S., and Patil, A. 2006. "Common Method Variance in IS Research: A Comparison of Alternative Approaches and a Reanalysis of Past Research," *Management Science* (52:12), pp. 1865–1883.
Pavlou, P. A. 2003. "Consumer Acceptance of Electronic Commerce: Integrating Trust and Risk with the Technology Acceptance Model," *International Journal of Electronic Commerce* (7:3), pp. 101–134.
Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., and Podsakoff, N. P. 2003. "Common Method Biases in Behavioral Research: A Critical Review of the Literature and Recommended Remedies.," *Journal of Applied Psychology* (88:5), pp. 879–903.
Ringle, C. M., Wende, S., and Becker, J. M. 2015. *SmartPLS 3*, Boenningstedt: SmartPLS GmbH, http://www.smartpls.com.
Rosen, L. D., Whaling, K., Carrier, L. M., Cheever, N. A., and Rokkum, J. 2013. "The Media and Technology Usage and Attitudes Scale: An Empirical Investigation," *Comput Human Behav*. (29:6), pp. 2501–2511.
Schmitz, C. 2015. *LimeSurvey Project Team*, LimeSurvey Project Hamburg, Germany, LimeSurvey: An Open Source survey tool.
Singh, T., and Hill, M. E. 2003. "Consumer Privacy and the Internet in Europe: A View from Germany," *Journal of Consumer Marketing* (20:7), pp. 634–651.
Söllner, M., Hoffmann, A., and Leimeister, J. M. 2016. "Why Different Trust Relationships Matter for Information Systems Users," *European Journal of Information Systems* (25:3), pp. 274–287.
Spiekermann, S. 2005. "The Desire for Privacy: Insights into the Views and Nature of the Early Adopters of Privacy Services," *International Journal of Technology and Human Interaction* (1:1), pp. 74–83.
The Tor Project. 2018. "Tor." (https://www.torproject.org, accessed February 20, 2018).
Venkatesh, V., and Davis, F. D. 2000. "A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Studies," *Management Science* (46:2), pp. 186–205.
Venkatesh, V., Thong, J., and Xu, X. 2012. "Consumer Acceptance and User of Information Technology: Extending the Unified Theory of Acceptance and Use of Technology," *MIS Quarterly* (36:1), pp. 157–178.

C.2 Anreize und Hemmnisse für die Implementierung von Privacy-Enhancing Technologies im Unternehmenskontext

David Harborth, Maren Braun, Akos Grosz, Sebastian Pape, and Kai Rannenberg. Anreize und Hemmnisse für die Implementierung von Privacy-Enhancing Technologies im Unternehmenskontext. In *Sicherheit 2018: Sicherheit, Schutz und Zuverlässigkeit, Beiträge der 9. Jahrestagung des Fachbereichs Sicherheit der Gesellschaft für Informatik e.V. (GI), 25.-27. April 2018, Konstanz*, pages 29–41, 2018. doi: 10.18420/sicherheit2018_02. URL https://doi.org/10.18420/sicherheit2018_02

H. Langweg, M. Meier, B.C. Witt, D. Reinhardt (Hrsg.): Sicherheit 2018,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2018 29

Anreize und Hemmnisse für die Implementierung von Privacy-Enhancing Technologies im Unternehmenskontext

Eine qualitative Analyse basierend auf Tiefeninterviews mit Privacyexperten

David Harborth[1], Maren Braun[1], Akos Grosz[1], Sebastian Pape[1], Kai Rannenberg[1]

Abstract: Wir untersuchen in diesem Artikel mögliche Anreize für Firmen Privacy-Enhancing Technologies (PETs) zu implementieren, und damit das Privatsphäre- und Datenschutzniveau von Endkonsumenten zu erhöhen. Ein Großteil aktueller Forschung zu Privatsphäre- und Datenschutz (im Weiteren *Privacy*) wird aktuell aus Nutzersicht, und nicht aus der Unternehmensperspektive geführt. Um diese bislang relativ unerforschte Lücke zu füllen, interviewten wir zehn Experten mit einem beruflichen Hintergrund zum Thema Privacy. Die Resultate unserer qualitativen Auswertung zeigen eine komplexe Anreizstruktur für Unternehmen im Umgang mit PETs. Durch das sukzessive Herausarbeiten zahlreicher Interdependenzen der gebildeten Kategorien leiten wir externe sowie unternehmens- und produktspezifische Anreize und Hemmnisse zur Implementierung von PETs in Firmen ab. Die gefundenen Ergebnisse präsentieren wir anschließend in einer Taxonomie. Unsere Ergebnisse haben relevante Implikationen für Organisationen und Gesetzgeber sowie die aktuelle Ausrichtung der Privacyforschung.

Keywords: Qualitative Tiefeninterviews; Qualitative Privacy Forschung; Privacy; Privacy-Enhancing Technologies; Firmenanreize

Also in dem Moment, wo ich sage: "Du hast hier Datenschutz und höhere Anonymität als Premium-Feature", dann hast du automatisch die Frage: „Ja, Standardkunden haben keinen Datenschutz bei euch?"

1 Einleitung

Privatsphäre- und Datenschutz (Privacy) stellen ein Grundrecht in der heutigen digitalisierten Welt dar (siehe dazu auch Datenschutz-Grundverordnung (DSGVO) der Europäischen Union [Re16]). Datenschutzfördernde Technologien (Privacy Enhancing Technologies, PETs), um diese auch umzusetzen gibt es bereits seit einigen Jahrzehnten. Allerdings werden PETs trotz technologischer Ausgereiftheit nur sehr vereinzelt verwendet [Fe01, Te17].
Dabei gibt es prinzipiell drei Akteure, die Anreize für eine entsprechende Verbreitung setzen könnten: Endverbraucher, Anbieter datenschutzbedürftiger Produkte oder Dienste

[1] Goethe University, Chair of Mobile Business & Multilateral Security, Theodor-W.-Adorno Platz 4, 60323 Frankfurt, Germany, david.harborth@m-chair.de

doi:10.18420/sicherheit2018_02

und Regulierer [Hi10, Xu12]. In bisherigen Studien wurde Privacy primär aus Perspektive der Endverbraucher untersucht [SDX11]. Ein Großteil der Endverbraucher räumt dabei anderen Faktoren als der informationellen Selbstbestimmung höhere Priorität ein. Dies zeigt sich beispielsweise an fehlender Zahlungsbereitschaft für Privacy [GA07] und daran, dass Faktoren wie Spaß die Privatsphärebedenken überlagern [DH06]. Rossnagel folgert auf Basis der Diffusionstheorie, dass Nutzer oft die Auswirkungen von PETs nicht erkennen können und deswegen für Anbieter die Vorteile der Einführung von PETs unklar sind [Ro10]. Marktwirtschaftliche Anreize, PETs einzusetzen wurden bisher für Anbieter nur in geringem Umfang untersucht. Rubinstein und der kanadische Datenschutzbeauftragte kommen dabei zum Schluss, dass aufgrund der niedrigen Nachfrage die marktwirtschaftlichen Anreize für Anbieter (oft privatwirtschaftliche Firmen) nicht groß genug sind und der Gesetzgeber Anreize schaffen sollte [Ru11, Te17]. Anreize fehlen möglicherweise auch deswegen, weil viele Geschäftsmodelle die Auswertung persönlicher Daten voraussetzen [Hu14]. Diese Strategie „verlässt“ sich zum Teil darauf, dass Anwender zu träge sind, Opt-out Optionen wahrzunehmen [Te17]. PETs, die Benutzern ein Opt-Out erleichtern würden, stehen dabei dem Geschäftsmodell entgegen.
Zusammengefasst zeigt sich, dass eine Erweiterung der Forschungsperspektive nötig ist. Die eher nutzerzentrierte Forschung muss durch Forschung aus Unternehmenssicht ergänzt werden. Es stellt sich daher die Forschungsfrage, welche Anreize und Hemmnisse Unternehmen dazu bringen bzw. davon abhalten, PETs in ihren Produkten zu etablieren.
Der Rest dieses Beitrags ist wie folgt aufgebaut: Kapitel 2 beschreibt den Forschungsstand und Kapitel 3 die verwendete Methodik. In Kapitel 4 stellen wir eine Taxonomie der Anreize und Hemmnisse für Firmen zur Einführung von PETs vor, die wir in Kapitel 5 diskutieren.

2 Aktueller Forschungsstand

Privacy-Enhancing Technologies stellt einen Sammelbegriff für verschiedene datenschutzfördernde Technologien dar. Borking und Raab definieren PETs als “coherent system of ICT measures that protects privacy [...] by eliminating or reducing personal data or by preventing unnecessary and/or undesired processing of personal data; all without losing the functionality of the data system” [BR01, S. 1]. Zusätzlich zu den PETs spielen sogenannte Transparency-Enhancing Technologies (TETs) eine wichtige Rolle dafür, dass Bürger bzw. Endverbraucher ihren Privatsphäre- und Datenschutz stärker wahrnehmen. TETs können folgendermaßen definiert werden: “[...] tools which can provide to the individual concerned clear visibility of aspects relevant to these data and the individual’s privacy” [Ha08, S. 205]. Zimmermann [Zi15] gibt einen ausführlichen Überblick am Markt existierender TETs. Die Unterschiede zwischen diesen Technologien sollen in diesem Beitrag nicht näher beschrieben werden, da es weitgehende Überlappungen zwischen ihnen gibt.
Privatsphäre- und Datenschutzthemen werden in bisherigen Studien primär aus Sicht des Individuums untersucht [SDX11]. Für unsere Forschungsfrage sind Studien interessant, die sich mit der Frage beschäftigen, inwieweit Individuen bereit sind, ihr Niveau an Privatsphäre- und Datenschutz zu erhöhen bzw. erhöhen zu lassen. Diese Fragestellung ist deshalb relevant,

da wir argumentieren, dass die Verantwortung für Privatsphäre- und Datenschutz von drei Parteien ausgehen kann, nämlich vom Individuum selbst, von Anbietern datenschutzbedürftiger Produkte oder von Regulierern. Die regulatorische Perspektive klammern wir in diesem Beitrag aus, da wir regulatorische Vorschriften mit möglichen Strafen bei Verstößen nicht als durch den Markt gegebenen Anreiz betrachten.
Forschung zu Privacy auf individueller Ebene hat gezeigt, dass Menschen angeben, sich um ihre Privatsphäre im Internet zu sorgen. Jedoch handeln sie dann entgegen ihrer vorherigen Aussagen und veröffentlichen beispielsweise zahlreiche persönliche Informationen in sozialen Netzwerken. Aktuelle Forschung erklärt dieses Verhalten einerseits mit einer Art kognitiver Dissonanz, die beim Thema Privacy hervortritt (vgl. Privacy Paradox [NHH07, SGB01]). Dieser Erklärung entgegenstehend sehen zahlreiche andere Forscher einen bewussten Trade-Off im Sinne eines Austausches eines speziellen Nutzens (kostenfreie Dienstleistung, Anerkennung, etc.) gegen Daten, auf den Nutzer sich einlassen (vgl. Privacy Calculus [DH06, DT15, DM16]). Weitere Forschung zeigt, dass Individuen neben dieser Divergenz von geäußerter Einstellung und beobachtbarer Handlung nur wenig Kosten (zeitlich und monetär) für ihre Privatsphäre tragen möchten [GA07].
Insbesondere der letzte Punkt wirft die Frage auf, inwieweit es unter diesen Voraussetzungen möglich ist, PETs profitabel am Markt zu etablieren. Daher ist es relevant, die Unternehmensperspektive auf Anreize für Unternehmen zu beleuchten. Die bisherige Forschung in diesem Gebiet ist nicht so reif wie im Gebiet der Forschung zu Privacy und Individuen [SDX11]. Einige der Artikel beschäftigen sich mit den Konsequenzen von Privatsphäre- und Datenschutzverletzungen in Firmen [AFT06] und wie Firmen mit diesen Verletzungen umgehen können [CKJ16]. Relativ viele Beiträge untersuchen, inwieweit Privacy ein kompetitiver Vorteil ist und sich in Geschäftsmodelle integrieren lässt [Ho14, CMHD15, Li11].

3 Methodik

In diesem Kapitel besprechen wir die verwendete qualitative Forschungsmethodik. Wir folgen diesem explorativen Ansatz, da bisherige Forschung unsere Forschungsfrage unzureichend adressiert hat. Im ersten Schritt haben wir einen semi-strukturierten Leitfadenfragebogen entworfen. Basierend auf dem semi-strukturierten Fragebogen werden die Teilnehmer durch das Interview geführt. Semi-strukturiert bedeutet in diesem Zusammenhang, dass das Interview maßgeblich durch die Interaktion und die Antworten des Befragten beeinflusst wird. Der Fragebogen hält nur besonders relevante Fragen fest, die auf jeden Fall angesprochen werden wollen. Dies hat den Vorteil, möglichst tiefe Einblicke und ausführliche Antworten vom Teilnehmer erhalten zu können. Der Fragebogen kann in drei inhaltliche Oberthemen aufgeteilt werden. Zuerst werden allgemeine Fragen zur Person und zum Unternehmen gestellt. Darauf folgen Fragen zu Privacy und PETs. Der zweite Teil deckt technische Fragen zum Status Quo und zu eventuellen zukünftigen Entwicklungen ab. Der dritte Teil behandelt ökonomische und gesellschaftliche Fragestellungen.
Für die Beantwortung unserer Forschungsfrage haben wir Experten und Professionals befragt, die mit Privacy-Enhancing Technologies (PETs) in ihren Unternehmen zu tun haben,

oder bei deren Produkten oder Dienstleistungen Privacy eine besondere Rolle spielt. Die Experten stammen von Firmen, die direkt PETs anbieten, oder in denen Privacy eine wichtige Rolle im Nutzenversprechen spielt. Als Beispiele hierfür sind der Telekommunikationssektor, Paymentprovider oder eCommerce Solutions Provider zu nennen. Wir haben zehn Interviews geführt und analysiert, wobei die Dauer zwischen 44 Minuten und 180 Minuten variiert. Die demografischen Informationen finden sich in Tabelle 2.
Die Interviews wurden alle aufgezeichnet und anschließend Wort für Wort transkribiert. Die Transkriptionen wurden daraufhin mit dem sog. offenen Kodieren und selektiven Kodieren analysiert [GS67, Ch14, St13]. Das offene Kodieren ist der erste Schritt der Datenauswertung und orientiert sich nah an den Daten (den Transkripten). Im nächsten Schritt werden Kodes zusammengefasst und abstrahiert (selektives Kodieren). Diese Schritte werden für jedes Interview einzeln durchgeführt und anschließend zwischen den Interviews. Diese sogenannte komparative Methode [GS67, Ch14, St13] ist ein elementarer Bestandteil der qualitativen Forschungsmethodik. Durch ständiges Vergleichen zwischen den Interviews, leiten wir abstrakte Kategorien aus den Daten ab, die ein vielfältiges Bild der Anreize und Hemmnisse liefert. Diese Kodierungsschritte wurden von zwei Autoren durchgeführt, um eventuelle Diskrepanzen in der Analyse der Daten festzustellen und zu lösen.

4 Resultate

Wir stellen in diesem Abschnitt das Kategoriensystem vor, welches elementar wichtig für eine logische und aufeinander aufbauende Strukturierung der Ergebnisse ist. Unsere übergeordnete Zielsetzung lag bei diesem Prozess darin, die Unternehmensperspektive in Bezug auf PETs nachzuvollziehen und zu verstehen. Da die Interviewteilnehmer sehr vielschichtige, sowohl vergleichbare und aufeinander aufbauende, als auch gegensätzliche, Stellungen bezogen, leitete sich hieraus eine argumentative Gliederung der relevanten Themenkomplexe ab. Diverse Querbezüge und Wechselwirkungen markieren damit ein interdependentes Gefüge, in welchem Unternehmen Anreize und Hemmnisse für die Implementierung von PETs betrachten. Ein Interviewteilnehmer fasst diese komplexen unternehmenszentrierten Abwägungsentscheidungen im Rahmen des Einsatzes von Privacy-Enhancing Technologies in den folgenden Worten zusammen: "*Ja, es [meint: PETs] soll funktionieren. Ja, und die, die es betreiben, sollen davon leben können. Ja, das soll so sein, aber es soll so sein, dass eben die Kontrolle, Transparenz und Nutzbarkeit breit akzeptierbar ist*" (D).

4.1 Technische Optimierung

Der Großteil der Interviewpartner gab an, dass PETs dienlich sind, um allgemein Unternehmensprozesse auf technischer und organisatorischer Ebene zu optimieren, "*dass man eben vor allem einen technologischen Vorsprung hat*" (B). Die spezifische Modellierung und Funktionalität der Technologie fördert dabei, dass Abläufe im Unternehmen unterstützt, vereinfacht und auch bedarfsgerecht angepasst werden können, was einen technologischen Ansatz zur Realisierung von Privacy-Maßnahmen im Unternehmen darstellt.

Tab. 1: Taxonomie

1. Technische Optimierung	1. Integration in den Geschäftsprozess	
	2. Datenmanagement und -vermeidung	
2. Geschäftsmodell	1. Weiterentwicklung Services	1. Kundenanforderung
		2. Vereinfachung / Convenience
		3. Awareness / Visualisierung
	2. Erweiterung Kundenkreis	1. Kerngruppe mit Privacybedürfnis
		2. Massenmarkt und Segmentierung
	3. Entwicklung neuer Geschäftsmodelle	1. Premiumservice
		2. Wirtschaftlichkeitsabwägung
	4. Positionierung für die Zukunft	
3. Unternehmens-wahrnehmung	1. (Technische) Sicherheit	
	2. Profilierung durch PETs	
	3. Geschäftsethik	

Integration in den Geschäftsprozess. Notwendige Bedingung für die Erwägung einer PET-Implementierung in den Geschäftsprozess ist, inwiefern Tools auf technischer Ebene auf relevante Prozesse abbildbar sind bzw. ob der Kostenaufwand in einem für angemessen erachteten Rahmen liegt: "*Gibt es so etwas? Wieso brauche ich so etwas? Wie kriege ich so etwas? Wie installiere ich so etwas? Und dann, wie setze ich es richtig für meinen Gebrauch ein?*" (H). Das Fehlen einer gesicherten Informationsgrundlage diesbezüglich wurde allerdings bemängelt: "*Man kann sich Vieles vorstellen, ob es dann in der Realität umsetzbar ist, ist dann die Frage*" (A).

Datenmanagement und -vermeidung. PETs können ein vereinfachtes und adäquates Datenmanagement gewährleisten, um darüber hinaus auch für den Geschäftsprozess nicht notwendige Daten vermeiden zu können. Dies kann letztlich auch mit dem gänzlichen Verzicht personenbezogener Daten einhergehen. Ein zentraler Anreiz für die Implementierung von PETs ist daher, dass Unternehmen die unmittelbare Entscheidungshoheit über Erhebung und Aggregation von Daten erlangen. Ein Interviewpartner erörtert, dass "*man immer von irgendwelchen Daten irgendwelche Rückschlüsse ziehen kann*" (A), weshalb Daten außerhalb ihrer jeweiligen Nutzung und Notwendigkeit als ein zusätzliches Geschäftsrisiko bewertbar sind. Ein weiterer Interviewpartner leitet aus der Vermeidung von Daten einen positiven Nutzen für Unternehmen ab: "*Wenn die Daten zum Beispiel nur dort sind, wo sie überhaupt gebraucht werden, dann brauche ich da nicht irgendwie auf den anderen Systemen, wo sie nicht gebraucht werden, erstmal Verschlüsselungen und Maßnahmen [mit]ergreifen*" (F). Die Möglichkeit auf schlankere und einfachere Unternehmensabläufe wurde ferner ebenfalls herausgearbeitet. Andererseits bieten unverschlüsselte personenbezogene Daten den Vorteil, eindeutig einem angelegten Profil und jeweiligen geschäftlichen Aktivitäten zugeordnet werden zu können: "*[Ein Klarname] ist natürlich einfacher, um Transaktionen zuzuordnen, um Versand beispielsweise einem gewissen Kunden zuzuordnen. Aber grundsätzlich wäre das auch über ein Pseudonym schon machbar*" (B).

4.2 Geschäftsmodell

Die Kategorie Geschäftsmodelle stellte sich im Rahmen unserer Auswertung als umfangreichste Kategorie dar (Schlüsselkategorie). Auf dieser Ebene wurden sowohl die stärksten Anreize als auch Hindernisse an die Forschenden herangetragen. Die zahlreichen Freiheitsgrade und Gestaltungsoptionen in Zusammenhang mit PETs wurden als primär ausschlaggebend für die hohe Schwingungsbreite des erwarteten Geschäftserfolgs bewertet. Der vorliegende Status quo wurde von einem Interviewpartner durchaus auch optimistisch aufgefasst: "*Wir können mit [PETs] glaube ich völlig neue Geschäftsmodelle aufbauen, die der Markt bisher überhaupt noch nicht kennt. [...] Wir wollen tatsächlich ein paar Schuhe von A nach B bringen anonym. Das kann aber auch was völlig anderes sein*" (E).

Weiterentwicklung Services. Unsere Ergebnisse zeigen, dass die Implementierung von PETs auch dazu beitragen kann, dass Unternehmen bestehende Services weiter in Richtung Datenschutz entwickeln können und damit in spezifischen Marktstrukturen einen Wettbewerbsvorteil generieren können.

KUNDENANFORDERUNG. Wie stark Kundenanforderungen hinsichtlich des Privatsphäre- und Datenschutzniveaus ausgeprägt sind, steht in Verbindung damit, welche Kundenstruktur gegeben ist und auf welche Segmente künftig spekuliert wird. Aus den Interviews ging sowohl hervor, dass es Kunden gibt, die ein großes Interesse hegen, sich zu schützen und sich diesbezüglich mit Nachfragen wie Ansprachen an Unternehmen wenden, als auch die Auffassung der Interviewpartner, dass Privacy keine bzw. eine eher untergeordnete Rolle beim Gros der (potenziellen) Kunden spielt. Eine etwas andere Konnotation sehen wir indes darin, dass Kunden einen ausreichenden Privatsphäre- und Datenschutz selbstredend erwarten, dies allerdings nicht zwingend explizit an Unternehmen herangetragen: "*Da erwartet der Kunde auch, dass da gewisse Schutzmechanismen passieren. [...] Und das passiert auch. Und das bezahlen sie auch impliziť*" (D).

VEREINFACHUNG UND CONVENIENCE. "*Jetzt speziell auf das Internet gesehen, [. . .] jeder gibt irgendwie Daten dort preis. [...] Ist halt oft schwer, nur das preiszugeben, was man möchte, weil man eben doch oft Dinge preisgibt, von denen man nicht weiß oder in dem Moment, wo man sie preisgibt, eben nicht weiß, was damit geschieht letztendlich*" (B). Im Anschluss an diesen Problemaufriss fassen wir unter dieser Kategorie, dass (mögliche) Kunden unbefangen eine Geschäftsbeziehung mit einem Anbieter eingehen und aufrechterhalten können, da entsprechende Services hinsichtlich ihrer Privacy-Einstellungen verbessert wurden. Den Konsumenten wird dadurch kommuniziert, dass sensible Daten nicht erhoben bzw. diese ausreichend durch jeweilige Mechanismen geschützt werden: "*[Die Kundenansprache] könnte man jetzt so machen: '[...] Wir haben jetzt eine neue Technologie [...] installiert [...] und die Möglichkeit schützt deine privaten Daten. Sonst bleibt für dich alles gleich.' [...] Es ist leicht verständlich. Er muss die Technologie auch nicht verstehen*" (A). Andererseits betrachtete ein Teil der Interviewpartner Einfachheit und Bequemlichkeit unter dem Gesichtspunkt, dass datenschutzfördernde Tools aus ihrer Sicht eher einen Mehraufwand für Kunden darstellen: "*Letztendlich, warum wir die Daten speichern möchten oder teilweise speichern wollen, ist eben, um dem Kunden zu vereinfachen, dass er beim nächsten Mal zum Beispiel dann nichts mehr eingeben muss. [...] Also letztendlich ist*

das immer so eine Abwägung zwischen Privatsphäre und zu viele Daten sammeln oder Einfachheit, also im Grunde eben ein möglichst einfaches Interface zu bieten" (B).

AWARENESS UND VISUALISIERUNG. Unternehmen wurde eine wichtige Rolle dabei zugesprochen, auf die Privacy-Thematik aufmerksam zu machen und (potenzielle) Kunden hierfür zu sensibilisieren. Die Interviewteilnehmer erachteten dies als angemessene Maßnahme, um Nachfrage auf diesem Gebiet zu generieren. Gleichzeitig sah ein Befragter Firmen in dieser Hinsicht nicht in der Verantwortung: "*Wir brauchen Awareness von den verschiedenen Segmenten, die es brauchen*" (H). Weiterhin wurde eine geeignete Form der Visualisierung als substanziell eingestuft, um Nutzern die Vorteilhaftigkeit von PETs vor Augen zu führen: "*Irgendwo hätte ich schon gerne als Endteilnehmer, wenn ich schon bezahle, ja, warum bezahle ich eigentlich? Also da muss irgendwie so eine Beweisnotwendigkeit sein*" (D). Inwiefern ein Premium- oder Upselling-Preismodell als sinnvoll zu erachten ist und ob ein Zusatznutzen wie die genannte Visualisierung einen attraktiven Trade-off für Kunden darstellt, werden wir im Abschnitt "Premiumservice" näher diskutieren.

Erweiterung Kundenkreis. Eine Privacy-freundliche Ausrichtung von Unternehmen kann neue Kundenmärkte öffnen und ein Alleinstellungsmerkmal darstellen.

KERNGRUPPE MIT PRIVACY-BEDÜRFNIS. Die Erweiterung des Kundenkreises ist ein häufig von den Befragten formulierter Anreiz für Unternehmen, PETs zu implementieren. Im Kern fielen darunter Personen, bei denen ein Privacy-Bedürfnis bereits überdurchschnittlich stark ausgeprägt ist: "*Ich adressiere genau diese Lücke. Ich adressiere die Freaks, ich adressiere die Nerds, ich adressiere diejenigen, die mehr Privacy Awareness haben, als die anderen*" (G). Nicht nur technikaffine und -interessierte Privatpersonen sind für die Befragten Teil dieser Kategorie, sondern auch Forschungs- und Entwicklungszentren sowie bestimmte Unternehmen. Neben intrinsischen Motiven PETs nachzufragen, spielen für Geschäftskunden häufig auch rechtliche Vorgaben zum Datenschutz eine zentrale Rolle.

MASSENMARKT UND SEGMENTIERUNG. Interviewteilnehmer lieferten sehr nuancierte Aussagen, bezüglich der möglichen Eignung von PETs im Massenmarkt (Primär- und Sekundärnutzen betrachtend). Wir haben diese vielfältigen Blickwinkel aufgegriffen, da sie im Rahmen ihrer jeweiligen Logik nicht zwingend als Widerspruch zu betrachten sind. Ein Potenzial zum Massenmarkt wurde unter anderem beschrieben, um ein besonders hohes Privatsphäre- und Datenschutzniveau als wünschenswerten Idealzustand in den Vordergrund zu stellen. Ein Befragter gab in diesem Zusammenhang an, dass "*jeder, der eine Kundenbeziehung hat*" (D), PETs im Sinne seiner Kunden implementieren sollte. Es ließ sich zudem der Konsens herauslesen, dass dies vom Großteil der Nutzer nicht explizit gefordert und nachgefragt wird, sondern eher akzeptiert, dann aber auch als positiver Nutzen empfunden wird: "'*Kann ich da Datenschutz einschalten? Ja oder nein?' Und dann glaube ich schon, dass viele Leute sagen: 'Joa, einschalten. Datenschutz ist immer gut'*" (G). Des Weiteren wurde Massentauglichkeit darin gesehen, dass PETs in bereits bestehende Produkte als Sekundärnutzen implementiert werden: "*Welche PETs setzen sich bisher im Massenmarkt durch? Nur eigentlich in Begleitung mit anderen Services eben*" (C). Als integraler Baustein etablierter Leistungen können PETs dadurch gar ohne aktives Nutzereinverständnis und ohne konkrete Nachfrage auf dem Massenmarkt etabliert werden. Gleichzeitig wurde die Notwendigkeit angeführt, in Marktsegmente zu unterscheiden: "*Seit Jahren habe ich*

gesagt, argumentiert, dass Mass Marketing ein Fehler ist. [...] Die Unterschiede zwischen den Anforderungen von den verschiedenen Segmenten [...] sind so groß. Verschiedene Leute brauchen unterschiedliche Unterstützung" (H). Diese Aussage steht in Verbindung damit, dass im Rahmen der geführten Interviews zahlreiche potenzielle Nutzergruppen genannt wurden: Unternehmen verschiedener Größe, Forschungsinstitutionen, öffentliche Einrichtungen, Privatpersonen, für die der Schutz der Privats- und Intimsphäre von hoher Bedeutung ist, beispielsweise, wenn sie "*in einem speziellen Segment sensible Produkte*" (A) erwerben möchten. Eine dritte Gruppe von Befragten bewertete einen größeren Kundenkreis hingegen als unrealistisch: "*Ich denke, dass es [meint: PETs] in gewisser Weise schon auch ein Nischenmarkt ist, also dass es nicht unbedingt massenmarkttauglich ist, weil einfach zu vielen Menschen die Privatsphäre da zu unwichtig ist [. . .] beziehungsweise unwichtig genug, um keine extra Mühen auf sich zu nehmen*" (B).

Entwicklung neuer Geschäftsmodelle. Neben der Erschließung neuer Kundenmärkte, kann eine datenschutzfreundlichere Ausrichtung neue Geschäftsmodelle ermöglichen.

PREMIUMSERVICE. Die Interviewten hatten keine eindeutige Meinung, inwiefern sich durch die Implementierung einer PET ein Premium- oder Upselling-Service geschäftlich sinnvoll ist. Ein Befürworter dieses Preismodells erklärte: "*Ja, es [meint: PETs] kostet was. Das ist wieder der berühmte Punkt: Es gibt etwas kostenlos, dann ist es aber eine mildtätige Spende, wo jemand sagt: 'Jawohl, ich spende das dafür, dass es auch wirklich kostenlos ist, so.' Alle anderen Sachen haben irgendwo ihren Trade-off. [. . .] Wir können genau sagen: 'Das kostet es, das bringen wir. Macht mit oder lasst es bleiben'*" (D). Allerdings können bestehende Leistungen des Unternehmens, die eventuell um keine datenschutzfördernde Technologie erweitert wurden, degradiert werden: "*Du versuchst ein Premium-Feature zu positionieren, aber gleichzeitig qualifizierst du alle anderen [angebotenen Services] ab und bringst die in eine Situation, dass du dich für die dann rechtfertigen musst: 'Warum kriegen das nicht alle?' Und die zweite Frage ist: Wer ist bereit für ein solches Premium-Feature zu bezahlen? Es ist dann irgendwas Exklusives*" (G). Daran anknüpfend bezieht ein Interviewteilnehmer folgende Position: "*Die monetären Kosten muss [das Unternehmen] kalkulieren*" (F).

WIRTSCHAFTLICHKEITSABWÄGUNG. Ohne Premium-Services müssen Unternehmen die Kostendeckung einer PET-Implementierung anderweitig garantieren, zum Beispiel durch Absatzsteigerung: "*Wir würden nur über Mengensteigerungen verdienen, weil wir dieses System anbieten*" (A). Alternativ ist es auch möglich, die Konversionsrate durch PETs zu steigern: "*Dem Hersteller nutzt es dann, wenn die Kunden einen Nutzen dahinter sehen und wenn es vielleicht dieser winzige Ausschlag ist, der eine Kaufentscheidung beeinflusst*" (G). Neben diesen quantifizierbaren Aspekten spielen weitere Faktoren, z.B. eine heuristisch orientierte Kosten-Nutzen-Analyse, eine zentrale Rolle in den jeweiligen Abwägungsentscheidungen. Diese gehen mit einer negativen Konnotation von Datenerhebungsvermeidung einher, bspw. durch Betrugsfälle: "*Ich denke [Unternehmen] werden auf jeden Fall erstmal Vorbehalte gegen so etwas [meint: PETs] haben, eben dadurch, dass sie befürchten, für irgendwelche Betrugsfälle oder so keinen greifbaren Kontakt irgendwie zu haben*" (B). Zum anderen wurde mehrfach folgende Befürchtung akzentuiert: "*Die [Unternehmen] haben natürlich kein Interesse an einem Pseudonym, denn die wollen ja Daten, Profile, Bewegungsprofile erstellen, weil das bares Geld ist*" (I).

Positionierung für die Zukunft. Zum einen betonten Befragte die Möglichkeit des Alleinstellungsmerkmales von PETs: "*Das wäre das Alleinstellungsmerkmal irgendwie für uns auch, [ein Produkt] eben anzubieten, [das] die Identität des Kunden schützt, was es eben zurzeit noch nicht so gibt, ja, also vor allem eben auch irgendwie dadurch einen Wettbewerbsvorteil zu gewinnen*" (B). Allerdings schwindet dieser Vorteil eventuell, wenn eine kritische Masse an Wettbewerbern ebenfalls vermehrt PETs implementieren oder Wettbewerber mit bedeutenderer Marktmacht bestimmte Schutztechnologien als neuen „Standard" etablieren: "*Die [meint: https-Verschlüsselung] setzt sich durch, langsam, weil tatsächlich große Konzerne auch dahinter stehen und das jetzt auch forcieren*" (C). Zum anderen wurde PETs eine präventive Wirkung beigemessen, um „Datenschutzskandale" zu vermeiden: "*Man [hat] das schon noch natürlich immer im Hinterkopf, weil man irgendwie auch ganz sicher nicht das Unternehmen sein möchte, was irgendwie in den Schlagzeilen ist, jetzt irgendwie auffällt dadurch, das die Privatsphäre nicht schützt.*" (B).

4.3 Unternehmenswahrnehmung

Privatsphäre- und Datenschutztechnologien verfügen über das Potenzial, sowohl die externe als auch die interne Wahrnehmung des Unternehmens zu beeinflussen.

(Technische) Sicherheit, Vertrauen und Qualität. Für das Vertrauensverhältnis zwischen Geschäftspartnern spielt das Verständnis der jeweiligen Technologie nur eine sekundäre Rolle. Die positive Wahrnehmung entstammt vorranging der impliziten Gefühlsebene: "*Heute verkaufen sich Sachen gut [...] indem gesagt wird: 'Wir machen das nach deutschen Datenschutzrechten [...].' Das verstehen die Leute. Die kennen überhaupt null Details dazu, aber die sagen sich: 'Okay. Wenn das nach deutschem Datenschutzding ist, dann passt das'*" (E). Die durch PETs gewährleistete Vertrauensfestigung kann sich dabei positiv auf den Ruf des Unternehmens niederschlagen: "*Ich sehe es als Qualitätsmerkmal*" (E).

Profilierung durch PETs. Die Kopplung von PETs an das bekannte Dienstangebot des Unternehmens stellt zudem ein kommunizierbares Alleinstellungsmerkmal dar, wie einer der Befragten erläuterte: "*Man kann es als Werbezweck verwenden. [...] Ich unterscheide mich damit von anderen. Das muss jetzt nicht sein, dass das so einen wahnsinnigen Zusatznutzen hat, es ist einfach ein Marketing-Effekt, den ich damit verbinden kann*" (C). Dieser allgemeine Werbeeffekt fördert dann nicht nur das reguläre Angebot, sondern auch die Reputation des Unternehmens: "*Ich glaube du kannst es [meint: mit PET-Implementierung auch Profitabilität sichern] nur machen, wenn du das als Add-on zu deinem Produkt [anbietest]. [...] Dann sagst du: 'Okay, ich investiere jetzt halt mal, weil das bringt mir vielleicht etwas in meinem Ansehen, in meinem Ruf, in meiner Zahl der [Kunden]'*" (I).

Geschäftsethik. Drei ethische Momente des unternehmerischen Handelns heben sich im Hinblick auf Privatsphäre- und Datenschutztechnologien aus den Interviews hervor. Erstens wird die These angeführt, dass Technologien und ihre Nutzung unabhängig von moralischen Wertepositionen gegeben sind: "*Es existiert, es ist keine Frage, keine moralische Frage. Es gibt Anonymität, das ist ein Konzept und es kann für verschiedene Zwecke benutzt werden*" (H). Diese an sich neutrale Auffassung von PETs kann polarisierenden Darstellungen gegenübergestellt werden. So können sich daraus moralisch vertretbare

Schritte zur informativen Sensibilisierung ergeben, jedoch auch verwerfliche, wie zum Beispiel eine einseitige, überspitzte Beängstigungskampagne. Diese können sich gar als geschäftsschädlich herausstellen: "*Natürlich, ich nutze es, ich habe Angst, aber ich weiß auch, dass ich das Produkt nur nutze, weil ich Angst habe. Es macht es jetzt auch nicht unbedingt so wahnsinnig sympathisch*" (C). Letztlich stehen moralische Aspekte der ökonomischen "Rationalität" von Firmen entgegen: "*Ich investiere jetzt in etwas und [...] ich mache das erst einmal, weil ich der Meinung bin: 'Das ist richtig und es hilft und es ist das Richtige zu tun und langfristig profitiere ich vielleicht auch davon, vielleicht nicht finanziell.' Das macht kein Unternehmen*" (I).

5 Diskussion und Schluss

Basierend auf der qualitativen Auswertung von zehn Tiefeninterviews mit Privacyexperten haben wir eine Taxonomie der Anreize und Hemmnisse für die Implementierung von PETs im Unternehmenskontext entwickelt.

Gemäß der Taxonomie spielen die mit Geschäftsmodellen verbundenen Anreize eine wichtige Rolle. Wie bestehende Literatur kommen wir allerdings zum Schluss, dass es Bedarf für weitere Forschung in dem Bereich zu Privatsphäre- und Datenschutz speziell im Unternehmenskontext gibt. Beispielsweise argumentiert Rubinstein [Ru11], dass die marktwirtschaftlichen Anreize für Firmen nicht gross genug sind und eine flächendeckende Verbreitung von PETs nur aufgrund von Initiativen des Gesetzgebers stattfinden wird. Ein weiteres vielversprechendes Thema für zukünftige Forschung besteht in dem Vergleich von Evaluierungen und Meinungen verschiedener Privacyexperten. Unsere Ergebnisse zeigen in einigen Bereichen kein klares Bild, da die Aussagen teilweise weit auseinander gehen. Befragte haben einerseits sehr unterschiedliche berufliche (Unterschiede in Firmen bezüglich Marktumfeld und Marktgröße) und private Hintergründe und andererseits sind ihre Positionen entweder ethisch oder praxisorientiert.

Wir tragen zur aktuellen Privacyforschung auf drei Wegen bei. Erstens haben wir Privacy im Unternehmenskontext, und nicht auf individueller Ebene, untersucht [SDX11]. Zweitens haben wir eine empirische, nicht normative, Studie durchgeführt, die auf einem Sample mit deutschen Interviewteilnehmern basiert. Zum Großteil ist Privacyforschung normativ und basiert auf Stichproben mit US-amerikanischen Teilnehmern [BC11]. Drittens haben wir mit einer qualitativen Methodik ein unterrepräsentiertes Thema explorativ von verschiedenen Dimensionen erforscht. Zusammenfassend zeigen unsere Ergebnisse, dass es durchaus Anreize für Unternehmen (abgesehen von Regulierung) geben kann, datenschützfördernde Technologien und Strukturen in ihren Geschäftspraktiken zu implementieren und damit dem Datenschutz zukünftig mehr Relevanz zu geben.

6 Acknowledgments

Diese Forschung wurde vom Bundesministerium für Bildung und Forschung (BMBF) unterstützt (Zuwendungsnummern 16KIS0371 and 16KIS0515).

Literaturverzeichnis

[AFT06] Acquisti, Alessandro; Friedman, Allan; Telang, Rahul: Is There a Cost to Privacy Breaches? An Event Study. In: International Conference on Information Systems (ICIS). 2006.

[BC11] Bélanger, France; Crossler, Robert E.: Privacy in the Digital Age: A Review of Information Privacy Research in Information Systems. MIS Quarterly, 35(4):1017–1041, 2011.

[BR01] Borking, John J.; Raab, Charles: Laws, PETs and Other Technologies for Privacy Protection. Journal of Information, Law and Technology, 1:1–14, 2001.

[Ch14] Charmaz, Kathy: Constructing Grounded Theory. Sage Publications, London, 2nd editio. Auflage, 2014.

[CKJ16] Choi, Ben C.F.; Kim, Sung S.; Jiang, Zhenhui (Jack): Influence of Firm's Recovery Endeavors upon Privacy Breach on Online Customer Behavior. Journal of Management Information Systems, 33(3):904–933, 2016.

[CMHD15] Casadesus-Masanell, Ramon; Hervas-Drane, Andres: Competing with Privacy. Management Science, 61(1):229–246, 2015.

[DH06] Dinev, Tamara; Hart, Paul: An extended privacy calculus model for e-commerce transactions. Information Systems Research, 17(1):61–80, 2006.

[DM16] Dienlin, Tobias; Metzger, Miriam J.: An Extended Privacy Calculus Model for SNSs: Analyzing Self-Disclosure and Self-Withdrawal in a Representative U.S. Sample. Journal of Computer-Mediated Communication, 21(5):368–383, 2016.

[DT15] Dienlin, Tobias; Trepte, Sabine: Is the privacy paradox a relic of the past? An in-depth analysis of privacy attitudes and privacy behaviors. European Journal of Social Psychology, 45(3):285–297, 2015.

[Fe01] Feigenbaum, Joan; Freedman, Michael J; Sander, Tomas; Shostack, Adam: Privacy engineering for digital rights management systems. In: Digital Rights Management Workshop. Jgg. 2320. Springer, S. 76–105, 2001.

[GA07] Grossklags, Jens; Acquisti, Alessandro: When 25 Cents is Too Much: An Experiment on Willingness-To-Sell and Willingness-To-Protect Personal Information. In: WEIS. 2007.

[GS67] Glaser, Barney G.; Strauss, Anselm L.: The Discovery of Grounded Theory. Aldine Pub., Chicago, 1967.

[Ha08] Hansen, Marit: Marrying Transparency Tools with User-Controlled Identity Management. In (Fischer-Hübner, S.; Duquenoy, P.; Zuccato, A.; Martucci, L., Hrsg.): The Future of Identity in the Information Society. IFIP — The International Federation for Information Processing, S. 199–220. Springer, Boston, MA, 2008.

[Hi10] Hirsch, Dennis D: The law and policy of online privacy: Regulation, self-regulation, or co-regulation. Seattle UL Rev., 34:439, 2010.

[Ho14] Hoffman, David: Privacy Is a Business Opportunity. Harvard Business Review, S. 2–5, 2014.

[Hu14] Hustinx, Peter: Preliminary Opinion of the European Data Protection Supervisor "Privacy and competitiveness in the age of big data: The interplay between data protection, competition law and consumer protection in the Digital Economy". Bericht, European Data Protection Supervisor, March 2014. available via https://edps.europa.eu/sites/edp/files/publication/14-03-26_competitition_law_big_data_en.pdf.

[Li11] Liu, Zhan; Bonazzi, Riccardo; Fritscher, Boris; Pigneur, Yves: Privacy-friendly business models for location-based mobile services. Journal of Theoretical and Applied Electronic Commerce Research, 6(2):90–107, 2011.

[NHH07] Norberg, Patricia A.; Horne, Daniel R.; Horne, David A.: The Privacy Paradox: Personal Information Disclosure Intentions versus Behaviors. Journal of Consumer Affairs, 41(1):100–126, jun 2007.

[Re16] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance). Official Journal of the European Union, L 119/1, http://data.europa.eu/eli/reg/2016/679/oj, April 2016.

[Ro10] Rossnagel, Heiko: The Market Failure of Anonymity Services. In (Samarati, Pierangela; Tunstall, Michael; Posegga, Joachim; Markantonakis, Konstantinos; Sauveron, Damien, Hrsg.): Information Security Theory and Practices: Security and Privacy of Pervasive Systems and Smart Devices: 4th IFIP WG 11.2 International Workshop, WISTP 2010, Passau, Germany, April 12-14, 2010, Proceedings. Springer, 2010.

[Ru11] Rubinstein, Ira S: Regulating privacy by design. Berkeley Technology Law Journal, 26(3):1409–1456, 2011.

[SDX11] Smith, H. Jeff; Dinev, Tamara; Xu, Heng: Theory and Review Information Privacy Research: An Interdisciplinary Review. MIS Quarterly, 35(4):989–1015, 2011.

[SGB01] Spiekermann, Sarah; Grossklags, Jens; Berendt, Bettina: E-privacy in 2nd generation E-commerce. In: Proceedings of the 3rd ACM conference on Electronic Commerce - EC '01. ACM Press, New York, New York, USA, S. 38–47, oct 2001.

[St13] Strübing, Jörg: 1978. 2013.

[Te17] Technology Analysis Division of the Office of the Privacy Commissioner of Canada: Privacy Enhancing Technologies - A Review of Tools and Techniques. Bericht, Office of the Privacy Commissioner of Canada, November 2017. available via https://www.priv.gc.ca/en/opc-actions-and-decisions/research/explore-privacy-research/2017/pet_201711/.

[Xu12] Xu, Heng; Teo, Hock-Hai; Tan, Bernard CY; Agarwal, Ritu: Effects of individual self-protection, industry self-regulation, and government regulation on privacy concerns: A study of location-based services. Information Systems Research, 23(4):1342–1363, 2012.

[Zi15] Zimmermann, Christian: A categorization of transparency-enhancing technologies. arXiv preprint arXiv:1507.04914, 2015.

All websites have been last accessed on December 5th, 2017.

A Demographische Daten der Interviewteilnehmer

Tab. 2: Demographische Daten der Interviewteilnehmer

Code	Branche	Unternehmensgröße Mitarbeiter	Unternehmensgröße Umsatz (in €)	Position	♂/♀	Dauer (hh:mm:ss)
A	Briefgesellschaft	1001-5000	50-100 Mio	Mitglied der Geschäftsleitung, Leiter Marketing und Vertrieb	♂	01:20:16
B	Zahlungssystem-Anbieter	51-200	n.a.	Produktionsmanager	♂	00:45:16
C	Energie-Beratung	11-50	1 Mio.	Geschäftsführer	♂	01:18:48
D	Anbieter E-Commerce-Lösungen	51-200	5-10 Mio.	Leiter Produktionsmanagement	♂	00:55:42
E	Anbieter E-Commerce-Lösungen	51-200	5-10 Mio.	Solutions Manager	♂	01:18:48
F	Anbieter E-Commerce-Lösungen	51-200	5-10 Mio.	Berater technische Pre-Sales	♂	00:44:57
G	Telekommunikation	0,1-0,5 Mio.	50-100 Mrd.	Experte Datenschutz-Audits und-Standards	♂	00:58:16
H	Telekommunikation	0,1-0,5 Mio.	50-100 Mrd.	Stellvertretender Leiter Datenschutz-Audits und-Standards	♂	00:58:16
I	Telekommunikation	0,1-0,5 Mio.	50-100 Mrd.	Leiter Datenschutz für Infrastrukturen und Dienstleistungen	♂	01:14:00
J	Beratung Technikfolgen-abschätzung IT	1-10	n.a.	Geschäftsführer	♂	01:51:26
K	Finanz-Dienstleister	50001-0,1 Mio.	20-50 Mrd.	Beraterin geschäftlicher Zahlungsverkehr	♀	01:49:48
L	Beratung IT-Management	1-10	n.a.	Geschäftsführer	♂	00:44:17

C.3 Towards an Architecture for Pseudonymous E-Commerce – Applying Privacy by Design to Online Shopping

Sebastian Pape, Daniel Tasche, Iulia Bastys, Akos Grosz, Joerg Laessig, and Kai Rannenberg. Towards an architecture for pseudonymous e-commerce – applying privacy by design to online shopping. In *Sicherheit 2018: Sicherheit, Schutz und Zuverlässigkeit, Beiträge der 9. Jahrestagung des Fachbereichs Sicherheit der Gesellschaft für Informatik e.V. (GI), 25.-27. April 2018, Konstanz*, pages 17–28, 2018. doi: 10.18420/sicherheit2018_01. URL https://doi.org/10.18420/sicherheit2018_01

H. Langweg, M. Meier, B.C. Witt, D. Reinhardt (Hrsg.): Sicherheit 2018,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2018 17

Towards an Architecture for Pseudonymous E-Commerce

Applying Privacy by Design to Online Shopping

Sebastian Pape[1] Daniel Tasche[2] Iulia Bastys[13] Akos Grosz[1] Jörg Lässig[2] Kai Rannenberg[1]

Abstract: In this paper we apply privacy by design in e-commerce. We outline the requirements of a privacy-aware online shopping platform that satisfies the principle of data minimization and we suggest several architectures for building such a platform. We then compare them according to four dimensions: privacy threats, transparency, usability and compatibility with existing business models. Based on the comparison, we aim to build the selected platform in the next step.

Keywords: privacy by design; pseudonymity; data minimisation; online shopping; e-commerce

1 Introduction

E-commerce is playing an increasingly important role for operators of shopping platforms and their customers. The estimated revenue for the German e-commerce market in 2017 amounts to €55 billion, while recent statistics forecast 58 million users and a market volume of €78 billion in 2021 [St16]. Shopping platform operators are collecting customer data, as personalized offers and recommendations lead to higher revenues. Despite an increase in public's awareness on the issue of data protection and growing concerns about the usage of their data, currently e-commerce users have no alternative to disclosing personal data and revealing shopping behavior [Jo16]. A recent study reveals that 50% of online services send full information about the users' baskets to Paypal (if PayPal was selected as payment method), which in turn forwards the information, *who* purchased *what* and *where*, to a third-party specialized in data aggregation [Pr16]. At least in Europe, these issues begin to be addressed through several regulations and directives. The General Data Protection Regulation (GDPR), planned to be applied in May 2018 in the EU countries, requires data protection by design and by default: "The controller shall implement appropriate technical and organisational measures, such as *pseudonymisation*, which are designed to implement data-protection principles, such as data *minimisation* [. . .] in order to [. . .] protect the rights of data subjects" [Re16]. Therefore, we aim to improve the processes in e-commerce in respect to the data protection principles pseudonymisation and data minimisation.
The e-shopping platform could track the user's online activity through IP address, third party web-tracking [MM12], browser fingerprinting [Ec10], canvas fingerprinting [Ac14],

[1] Goethe University, Chair of Mobile Business & Multilateral Security , firstname.lastname@m-chair.de
[2] University of Applied Sciences Zittau/Görlitz, {d.tasche,j.laessig}@hszg.de
[3] Chalmers University of Technology, Gothenburg, Swedenbastys@chalmers.se

doi:10.18420/sicherheit2018_01

or evercookies [Ac14]. We do not investigate them in this paper, as previous work has already suggested different countermeasures [DMS04, PCM13, Ba13, LRB16].
Following these requirements, we make a step forward in preserving customer's privacy in online shopping, by describing an e-commerce platform that satisfies the principles of data minimization and pseudonymization (Section 3). We then suggest several architectures for building such a platform (Section 4) and compare them based on the privacy threat analysis methodology LINDDUN [WJ15], but also with respect to usability, transparency and compatibility to existing business models (Section 5).

2 Related Work

Growing concern about user traceability when making electronic payments propelled efforts in the area of privacy-preserving e-commerce. Initial work mainly concentrates on anonymous electronic payment methods through cryptographic mechanisms such as blind signatures [Ch83, Ch85, CFN90]. Aiello et al. [AIR01] describe a cryptographic protocol for anonymous shopping of digital goods based on priced oblivious-transfer and private information retrieval [Ch95]. In their setting, the customer makes an initial deposit which is later used to retrieve the desired items. Besides the initial deposit and the interaction with the platform, the online shop learns nothing else. In particular, it does not learn what or how much is purchased, nor when the buyer runs out of credit. While interesting, this approach is not feasible for deployment, as the customer would have to download the entire encrypted database. More recent work brings several improvements to the underlying protocols [RR01, CDN09, CDN10, HOG11], but they still only focus on *digital* goods, while our interest is in achieving customer privacy when purchasing *physical* goods.
A first step towards anonymous and pseudonymous e-commerce addresses the problem of purchasing goods with digital assets in a privacy-friendly manner [Sa14, GGM16, Go17]. Goldfeder et al. [Go17] introduce a series of escrow protocols to use when buying physical products online and paying with Bitcoin. While some of these protocols satisfy strong security properties, the buyer is still required to provide the seller with an address for delivering the goods, breaking to some extent buyer anonymity. Even though the seller does not learn the exact address of the buyer (as the address of a friend or of a post office can be provided instead), the seller learns the location where the product has to be dispatched.

3 System Overview

First, we give a brief overview of the involved parties and the relevant data.

Involved parties. The system consists of the following five parties:
- The User is a (registered) customer interested in purchasing goods online from Shop.
- Shop is the party that sells the (physical) goods through a platform accessible via Internet.
- The payment provider Pay collects the payment from User and transfers it to Shop.
- The logistics provider Shipping delivers the purchased goods from Shop to User.
- ID-Provider is a third-party responsible for managing the user's profile.

In order to prevent the Shop from collecting customers' private data and creating dossiers that reveal shopping behavior, we introduce a trusted third party in the system, ID-Provider, that increases the usability and the privacy of the architecture. It is responsible with managing the User's real and generated identities. A customer registers with ID-Provider with the real identity, and receives from ID-Provider a new generated identity, a pseudonym for logging in with Shop. Basically it acts as an authentication provider with pseudonymous identities, single sign on system for online shops and shopping process management system that connects the stakeholder for one shopping procedure. ID-Provider increases the usability for the User as well as the privacy of the overall shopping process. We require the user to provide the real identity in order to prevent system abuse. The pseudonym can be lifted in case of proved misbehavior. User can use the same pseudonym on multiple online platforms, or can create several pseudonyms, one for every platform, or even one for every purchase on the same platform.

User data. For a successful purchase, the user needs to provide the following information:

- *Product data* refers to the products selected by User for purchase.
- *Total value* refers to the purchasing price of the selected products plus additional payment and shipping charges to User.
- *Payment data* represents the data needed for a successful payment. Depending on the selected payment method this can be name, full address, bank account or credit card number, or even an anonymous payment method as sketched in Section 2. In general, banks and financial service providers require more information about a payment than just the bank account and the total value.
- *Delivery data* represents the information the delivery service Shipping needs for a successful delivery to User. In most cases, this is the name and address of User. However, other options are container freight stations and poste restante delivery, which do not necessarily require the same information.

The identifiability of the User and the linkability of purchases by Pay and Shipping depends on the chosen payment and delivery methods and applies to all architecture scenarios we will further discuss (Section 4). We assume that none of the parties collude, as collusion between Shop and any of ID-Provider, Pay or Shipping is sufficient for User profiling.

System requirements

A representation of a current e-shopping process is depicted in Figure 1. In general, Shop collects the User's data required for payment processing and package delivery. While it is possible to use a payment provider, such as Paypal [Pa17], and not provide Shop with any payment information, in most cases, the payment provider offers Shop a possibility to manage the payments and allows it to access the user's payment data.

As already discussed in Section 1, we ignore other customer tracking possibilities and focus our analysis on the data provided by User to the other parties. If a privacy-friendly online shopping platform would exist, the users could try to protect themselves via technical measures or legislation could protect the users by banning tracking without their consent.

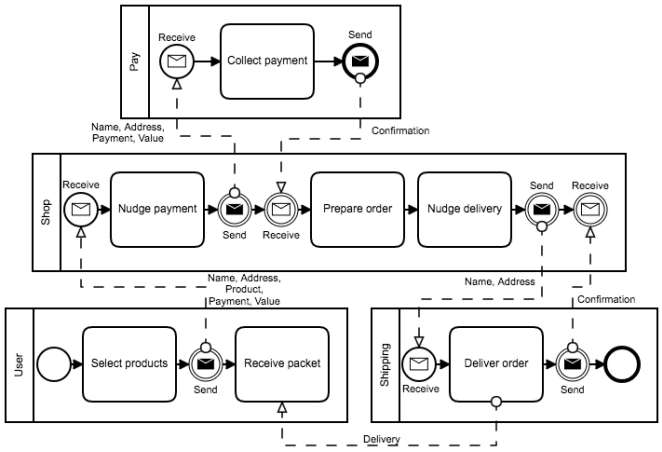


Fig. 1: Traditional Shopping Process in Business Process Modeling Notation [Ob11]

Since the login process will not differ much for the proposed architectures, the data in focus are product data, the value of the products, payment data and shipping data. When designing the pseudonymous e-commerce architecture, we aim for the *principle of minimum disclosure* under the constraints that the usability of the system should be comparable to the current systems in practice and that the process should be as transparent as possible to the user. Additionally, as discussed in Section 1, to promote a widespread use of our architecture, the shop providers business model should be respected. To chose our basic architecture, we consider the following dimensions for requirements and comparison:

Privacy. Shop should learn only the user's activity on the platform, i.e. the purchased products and their total value. The vendor does not learn payment or shipping data. Pay should learn nothing except for the amount to be payed by the user to the vendor and the payment data from the user. More specifically, the payment service does not learn the products the user purchased, but only their total value. Shipping should learn only the shipping data, but not the content (purchased products) of the package(s) to be delivered.

For the privacy analysis we apply LINDDUN, a privacy threat analysis methodology [De11, WJ15] which supports analysts in eliciting privacy requirements – similar to the security threat modeling framework STRIDE [Sh14]. Since we have a manageable number of entities in the context of our architecture scenarios, we don't run into the risk of threat explosion and can use the LINDDUN privacy analysis framework to systematically account for privacy specific threats. It is based on the graphical representation of the system's abstract representation by a data flow diagram (DFD) and the subsequent mapping of the framework's six high-level threats to each DFD element. Therefore, we model the process from checkout via payment to the delivery procedure of the products in a DFD. For each entity, we investigate the threats and map them to the elements in the DFD. In the following, we briefly discuss the privacy threats we are going to analyze.

Since the user should be able to shop pseudonymously, we consider the *identifiability* of the user (e.g. by payment or delivery data) as the main threat. In that context it is also important which parties hold which data. Depending on the party, information on *purchased products*, *the value of the purchased products*, *payment data* and *delivery data* is necessary for providing the service. As discussed, in particular the last data is suitable to revoke the User's pseudonymity. Therefore, we consider the *disclosure* of this *data* as another threat. Even if the User is not directly identifiable, *linkability* of two (or more) purchases of a user could reveal sensitive information leading to identification or at least the building of a meaningful profile. We further investigate which of the parties is able to *detect login*, *purchase*, *payment* and *delivery* events which could also be used to profile the User. Detectablility of one of the events does not mean the corresponding data is revealed, but in most cases the involved party can be identified (e.g. User did payment with Pay but neither amount nor payment data can be seen). *Unawareness* and *non-compliance* are out of the scope, as they are more related to the user interface and the entities' policies which are independent of our system architecture process. We are also not regarding *non-repudiation* for this paper since we consider it more related to contracting and legal aspects than to the architecture of our shopping platform.

Usability. Many aspects concerning the usability will not depend on the system's architecture, but on proper user interfaces allowing the user to manage his data in a easy and transparent way. However, in order to allow the user to easily use the system from different clients (e.g. computer, tablet, smart phone, . . .), the user should not store information such as a cryptographic key. Additionally, the speed of the system should be comparable to existing systems, thus complex cryptographic protocols which delay the process too much can not be used. As a consequence, certain privacy enhancing technologies such as attribute based credentials [SKR12] do not come into play, because they make use of cryptographic keys, which the user would have to store on a smartcard. We compare the different architectures based on the effort the user needs to take for.

Transparency. A natural data flow which allows the user to easily understand which data is provided to whom for which purpose contributes to a transparent system. Since the user interface is out of the scope of this work, transparency of the different architectures will depend only on data flows.

Compatibility to existing business models. Analogous to attribute based credentials [Sa15], we assume that when preserving the online shop providers' business models, a broad distribution of our platform can be more easily achieved. Certainly, this does not mean that the shop providers should be allowed to collect all data they want. But allowing them to keep profiles for pseudonyms and sending e.g. newsletters (via ID-Provider) to users who gave consent would certainly be helpful for the adoption of pseudonymous e-commerce.

4 Architecture Variants

In this section we describe the three architectures, we considered for implementation. For an easier comparison, we also analyzed the current shopping process. The standard architecture allows the Shop to gather a big volume of data about its users. In order to

avoid this, we suggest three architectures, two of them make use of public-key infrastructure (see Sections 4.2 and 4.3) and a third one without encryption but self data hosting (see Section 4.4). All scenarios involve an ID-Provider for managing the user's profile.

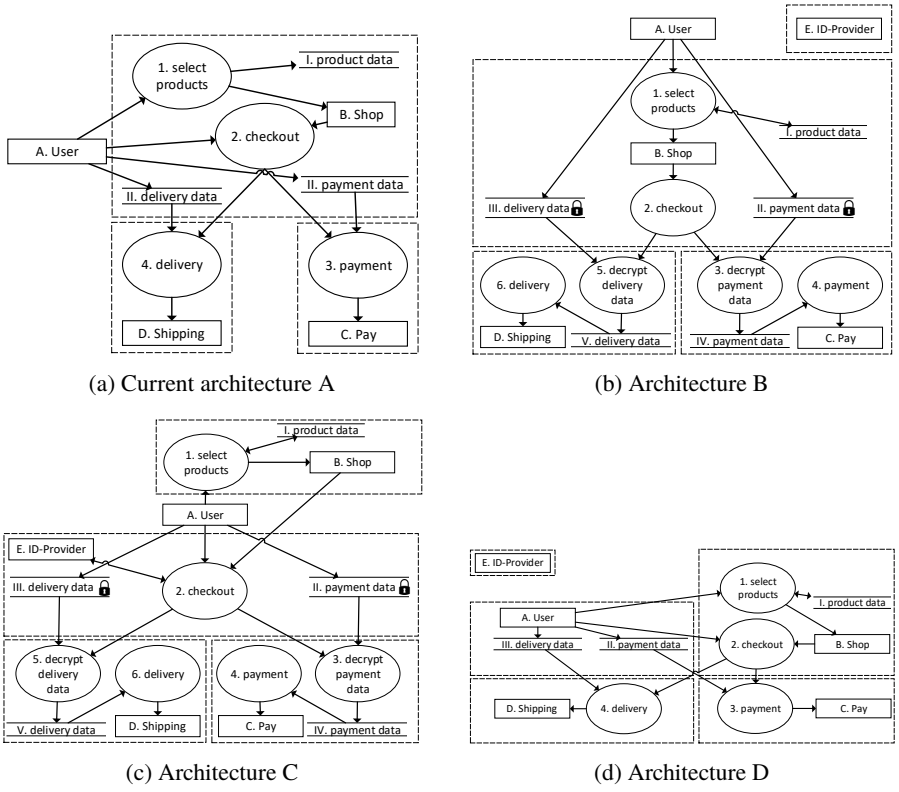
For the following analysis, we abstract from the login process and from confirmations as far as possible. Although other variants exists, we assume the User selects the products, pays and gets them delivered afterwards. Special care has to taken that Pay and Shipping providers do not pass the User's data to the Shop, e.g. by offering an administrative user interface, where payment data is listed or sending tracking information of the delivered packages. Each architecture's description follows the following template: We describe the process of every scenario and briefly discuss advantages and disadvantages. The corresponding data flow from selecting the products, checkout, payment and delivery process is depicted in Fig. 2. The analyzed privacy threats described in Section 3 are listed in Tab. 1. For each privacy threat (from Sect. 3) we denote the scenarios where it exists. For some of the analyzed threats, it depends on the users. If users don't want the shop to link their payments, they can use a new pseudonym for each purchase. For payment and shipping it depends on the kind of service the user chooses. Clearly, it makes a difference whether they are paying with anonymous electronic payment or by providing their credit card data. For shipping they could ask for home delivery or use a container freight station. We denote these threats in brackets in Tab. 1.

4.1 A: Current Shopping Process

The standard shopping process is depicted in BPMN in Figure 1 and has already been described. Figure 2a shows the data flow diagram. The Shop collects all information about the user, and thus can identify the user and can link all shopping activities. The identifiability of the User and the linkability of purchases by Pay and Shipping depends on the chosen payment or delivery method. The highest privacy threat for the User is the Shop because of the possibility to disclose the User's payment and delivery data as well as profiling the User.

4.2 B: Shop Stores Encrypted Data

In this scenario, the user reveals only his real identity (name) to ID-Provider when registering. The ID-Provider acts as single-sign-on login service, to allow the user to log in several Shops without further registration. Additionally, ID-Provider provides public keys for payment and shipping provider. Shipping and payment data is stored encrypted on the Shop's server. The data flow of this scenario is depicted in Figure 2b.
The User initiates the process by *select products*. The Shop gets the product data and stores it. In the *checkout* process, the User decides on a payment and shipping provider. The user gets the public keys for any provider he wants to use, encrypts the payment respectively delivery data and sends it to the Shop. Subsequently, the Shop initiates the *payment* process by forwarding the encrypted banking details along with the amount to be payed to Pay. After successfully decrypting the payment data and completing the payment transaction,



(a) Current architecture A (b) Architecture B

(c) Architecture C (d) Architecture D

Fig. 2: Data flow diagram

Pay sends a confirmation of payment to Shop. Upon confirmation, Shop starts the *delivery* process and sends the package labeled with the User's encrypted address to Shipping. After successfully decrypting the delivery data, Shipping proceeds with delivering the package and provides Shop with a confirmation of delivery. In this architecture Shop is not able to identify the user and can not disclose payment and delivery data. Since there is a possibility to use one pseudonym for each shopping process, the shop is also not able to link the purchases of an user unless the User allows it. The ID-Provider only knows the real identity of the user, but can not disclose payment or delivery information and also does not learn anything about the purchase process. However, the ID-Provider is able to detect the login process. The information distribution of Pay and Shipping are not affected. Therefore, the same conditions apply as for the standard architecture.

Advantages and disadvantages.

+ ID-Provider does not learn methods User uses for payment.
+ The Pay and Shipping services do not learn the virtual identity of User.
– The ID-Provider is able to detect logins in the store.
– Key management: It is difficult for the User to encrypt the payment and delivery data.

Tab. 1: Privacy Threats Mapped to Architecture Variants from Sect. 4

Threat \ Entity	Shop	Pay	Ship	Identity Provider
Identifiability	A	(ABCD)[2]	(ABCD)[3]	B C D
Disclosure shopping cart	A B C D			
Disclosure total value	A B C D	A B C D		
Disclosure payment data	A	A B C D		
Disclosure delivery data	A		A B C D	
Linkability purchase	A(BCD)[1]	(ABCD)[2]	(ABCD)[3]	C
Detectablility login	A B C D			B C D
Detectablility purchase	A B C D	A B C D	A B C D	C
Detectablility payment	A B C D	A B C D		C
Detectablility delivery	A B C D		A B C D	C

[1] Depends on the user's choice. [2] Depends on user's payment [3] Depends on user's shipping

Either this is done in the browser (e.g. with Javascript) or by an App, but the user has to trust the party providing the code.

4.3 C: ID-Provider Stores Encrypted Data

This architecture is similar to architecture B. The only change is that the (encrypted) payment and delivery data is stored at the ID-Provider. As a consequence, instead of directly delivering the data to Pay and Shipping, the Shop refers them to the ID-Provider where they need to authenticate and ask for the User's data. Therefore, the data flow itself is very similar to the one of architecture B (see 4.2) as depicted in Figure 2c. In this scenario, ID-Provider controls the shopping process. It knows the identity of the user and has information about where the user shops but does neither know the payment or delivery data since they are encrypted nor any details of the shopping content. Shop does not know the real identity of the user nor the payment or delivery details. The data distribution or possible disclosure from Pay and Shipping are unchanged.

Advantages and disadvantages.

+ ID-Provider does not learn the payment or delivery data of the User.
– If the User does not perform the encryption himself, then he has to trust ID-Provider to provide him with the correct public keys of the payment and logistics services.
– ID-Provider learns the Shop where the User makes his purchases.
– One more point of failure: ID-Provider is involved in multiple transactions.

4.4 D: User Gets Redirected to 3rd Parties

In this scenario, ID-Provider solely acts as a single-sign-on service and certification authority. All the information required for each of the steps of the pseudonymised shopping process is stored by the User. He initiates the process by *selecting the products*. The Shop gets

the product data and stores it. In the *checkout* process, the User decides on a payment and shipping provider. Subsequently, the Shop redirects the User for the *payment* process and the User delivers his payment data directly to Pay. Pay sends a confirmation of payment back to Shop. Upon confirmation, the Shop redirects the User for the *delivery* process and the User delivers his delivery data directly to Shipping. Shipping receives the package from Shop with an identifier to link it to the address and proceeds with delivering the package and provides Shop with a confirmation of delivery. Figure 2d shows the data flow. Since the User has full control about his profile data, Shop does neither know the User's identity nor the payment or delivery data. The information distribution of Pay and Shipping are not affected. Therefore, the same conditions apply as for the standard scenario.

Advantages and disadvantages.

+ The User is fully in control of his personal information.
+ Only necessary data is provided to each party.
+ Only communication needs to be encrypted.
– Additional tools have to be provided for Users to host their information.
– A lot of transactional load is put on the User. In particular, the User has to check that he is providing the information to the correct party, e.g. by checking cryptographic certificates.
– The payment process has to work instantly, otherwise additional communication is needed to synchronise payment with shipping processes.

5 Architecture comparison

In this section we compare the previously described architectures on the four dimensions described in Section 3: privacy, usability, transparency, and compatibility.

Privacy. Every involved party should learn only information about the activity belonging to its area of responsibility. In the standard scenario the Shop holds every information about the User's identity. As described in Sect. 4, all proposed architectures consider the principle of minimum disclosure. They differ only in the information provided to the ID-Provider.
In each scenario the User has the possibility to create several shopping pseudonyms. If he uses one for every shopping process, the store could not link several purchases. This applies to all architectures. The linkability of the purchase and identifiablity of the User on Pay's and Shipping's side depends on the payment and shipping methods and is not an architectural aspect. ID-Provider could link the purchases in Scenario C because ID-Provider manages the checkout process. In the other scenarios ID-Provider acts as a real identity provider and just manages the login process. Therefore, the detectability of a purchase, a payment and a delivery applies to Scenario C, but not to Scenario B and Scenario D.

Usability. Every architecture has the registration at ID-Provider and the managing of pseudonyms in common. That means compared to the standard scenario one has to maintain data not on Shop's side but on ID-Provider's side. As a compensation for managing the profiles, the User would not need to register at any Shop anymore.
Architecture B and C come with additional effort since the Users have to encrypt their data. In particular, in architecture B, Users face the problem that they might not want to

trust the Shop's App or Javascript-code making it difficult to encrypt. On the other hand in architecture C, the user has to register at the ID-Provider anyway and it seems reasonable to rely, e.g. on an App or Javascript-code on a web page. Architecture D asks the user to provide his payment and delivery data for each purchase again. This could be mitigated by making use of the The PaymentRequest API [Ba17]. However, since the recommendation is quite new, it will take some time until this has been adapted. For the authentication and single sign-on the X.509 standard could be used but needs some extensions to provide special user information. Therefore, Dash et al. [Da17] show an architecture proposal for an identity management architecture as a service. Additionally, since the Shop redirects the User to Pay and Shipping, the User has to check for each of the providers that Shop was directing her to the correct entity and not to a forged one to get the User's data.

Transparency. Despite sharing payment and delivery data directly with the Shop the standard architecture is quite transparent, because the User should be aware of sharing this data with the Shop. Although, the user might not be aware that this information might be shared with or is accessible by 3rd party service providers (e.g. webhoster, payment provider). The same holds for architecture D, where the Users need to provide their data to each entity directly. Architectures B and C, lack a bit of transparency, because it is harder for the user to assess how and from whom the encrypted data will be processed. However, it's up to the respective entity to inform the User in a supporting way.

Compatibility. The basic business processes of the involved parties are not broken by this architectures. However, by not disclosing the User's identity and therefore contact information to Shop, Shop needs to rely on ID-Provider to forward e.g. newsletters or special offers to the User. In case of misuse or disputes, ID-Provider is needed to reveal the User's identity. Pay and Shipping need to adjust their processes, in order to not reveal the User's data to Shop. However, there is no large difference here between architectures B, C, and D.

Final Architecture. Table 2 shows an overview of all attributes concerning the four analyzed architectures. While architectures B and D are favorable in respect to privacy and transparency, our focus when defining the requirements was to put emphasis on usability. Improved privacy should not complicate the shopping process for the user. The slight disadvantage in transparency from architecture C to D does not outweigh the disadvantage of architecture D that Users need to provide their data for each purchase again or alternatively have additional accounts (and logins) at payment and shipping providers. Therefore, we believe architecture C to be the most feasible option.

	Privacy	Usability	Transparency	Compatibility
A	-	o	+	++
B	++	+	o	+
C	+	++	o	+
D	++	o	+	+

Tab. 2: Comparison of the several architectures.

6 Conclusion and Future Work

In the context of pseudonymous online shopping, we presented and assessed three different architectures and compared them to the existing architecture. So far, the proof of concept shows, that a pseudonymous e-commerce process can be set up in a usable and privacy-friendly way. The User data is no longer on Shop's side but split to several parties that are involved in the shopping process.
We plan to add more processes to the shopping system such as returning goods and writing invoices. Around this, several legal and technical issues need to be resolved, e.g. how the Shop can issue an invoice to a pseudonym. Even though the PaymentRequest API only supports non-normative encryption of data fields and might also expose payments methods (cf. [Ba17, Section 19.2]), it might be helpful in storing payment and delivery data in the User's computer to avoid creating a centralized database.
Future work includes the implementation of certain restrictions for Users. For example, only Users above certain age or in certain geographical regions can access certain products. The next steps also contain the detailed description of the used protocol.

7 Acknowledgments

The SIOC project [SI] is supported by the German Federal Ministry of Education and Research's (BMBF) program "Datenschutz: selbstbestimmt in der digitalen Welt".

References

[Ac14] Acar, Gunes; Eubank, Christian; Englehardt, Steven; Juarez, Marc; Narayanan, Arvind; Diaz, Claudia: The web never forgets: Persistent tracking mechanisms in the wild. In: CCS. 2014.

[AIR01] Aiello, William; Ishai, Yuval; Reingold, Omer: Priced Oblivious Transfer: How to Sell Digital Goods. In: EUROCRYPT. 2001.

[Ba13] Bau, Jason; Mayer, Jonathan; Paskov, Hristo; Mitchell, John C: A promising direction for web tracking countermeasures. W2SP, 2013.

[Ba17] Bateman, Adrian; Koch, Zach; McElmurry, Roy; Denicola, Domenic; Cáceres, Marcos: , Payment Request API. https://www.w3.org/TR/2017/CR-payment-request-20170921/, 2017. W3C Candidate Recommendation 21 September 2017.

[CDN09] Camenisch, Jan; Dubovitskaya, Maria; Neven, Gregory: Oblivious transfer with access control. In: CCS. 2009.

[CDN10] Camenisch, Jan; Dubovitskaya, Maria; Neven, Gregory: Unlinkable priced oblivious transfer with rechargeable wallets. In: FC. 2010.

[CFN90] Chaum, David; Fiat, Amos; Naor, Moni: Untraceable electronic cash. In: CRYPTO. 1990.

[Ch83] Chaum, David: Blind signatures for untraceable payments. In: CRYPTO. 1983.

[Ch85] Chaum, David: Security without identification: Transaction systems to make big brother obsolete. CACM, 1985.

[Ch95] Chor, Benny; Goldreich, Oded; Kushilevitz, Eyal; Sudan, Madhu: Private information retrieval. In: FOCS. 1995.

[Da17] Dash, Pritam; Rabensteiner, Christof; Hörandner, Felix; Roth, Simon: Towards Privacy-Preserving and User-Centric Identity Management as a Service. In (Fritsch, Lothar;

Roßnagel, Heiko; Hühnlein, Detlef, eds): Open Identity Summit 2017. Gesellschaft für Informatik, Bonn, pp. 105–116, 2017.

[De11] Deng, Mina; Wuyts, Kim; Scandariato, Riccardo; Preneel, Bart; Joosen, Wouter: A privacy threat analysis framework: supporting the elicitation and fulfillment of privacy requirements. Requirements Engineering, 2011.

[DMS04] Dingledine, Roger; Mathewson, Nick; Syverson, Paul: Tor: The second-generation onion router. Technical report, Naval Research Lab Washington DC, 2004.

[Ec10] Eckersley, Peter: How unique is your web browser? In: PETS. 2010.

[GGM16] Garman, Christina; Green, Matthew; Miers, Ian: Accountable privacy for decentralized anonymous payments. In: FC. 2016.

[Go17] Goldfeder, Steven; Bonneau, Joseph; Gennaro, Rosario; Narayanan, Arvind: Escrow protocols for cryptocurrencies: How to buy physical goods using Bitcoin. 2017.

[HOG11] Henry, Ryan; Olumofin, Femi; Goldberg, Ian: Practical PIR for electronic commerce. In: CCS. 2011.

[Jo16] Jourova, Vera: , How does the data protection reform strengthen citiziens' rights? http://ec.europa.eu/justice/data-protection/document/review2012/factsheets/factsheet_dp_reform_citizens_rights_2016_en.pdf, 2016.

[LRB16] Laperdrix, Pierre; Rudametkin, Walter; Baudry, Benoit: Beauty and the beast: Diverting modern web browsers to build unique browser fingerprints. In: IEEE S&P. 2016.

[MM12] Mayer, Jonathan R; Mitchell, John C: Third-party web tracking: Policy and technology. In: IEEE S&P. pp. 413–427, 2012.

[Ob11] Object Management Group: , Notation (BPMN) version 2.0. OMG Specification, 2011.

[Pa17] Paypal: , Paypal Website. https://www.paypal.com, 2017.

[PCM13] Perry, Mike; Clark, Erinn; Murdoch, Steven: The design and implementation of the Tor Browser. Technical report, The Tor Project, 2013. https://www.torproject.org/projects/torbrowser/design/.

[Pr16] Preibusch, Sören; Peetz, Thomas; Acar, Gunes; Berendt, Bettina: Shopping for privacy: Purchase details leaked to PayPal. Electronic Commerce Research and Applications, 2016.

[Re16] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance). Official Journal of the European Union, L 119/1, http://data.europa.eu/eli/reg/2016/679/oj, 2016.

[RR01] Ray, Indrakshi; Ray, Indrajit: An Anomymous Fair Exchange E-commerce Protocol. In: IPDPS. 2001.

[Sa14] Sasson, Eli Ben; Chiesa, Alessandro; Garman, Christina; Green, Matthew; Miers, Ian; Tromer, Eran; Virza, Madars: Zerocash: Decentralized anonymous payments from bitcoin. In: IEEE S&P. 2014.

[Sa15] Sabouri, Ahmad: Understanding the Determinants of Privacy-ABC Technologies Adoption by Service Providers. In: Open and Big Data Management and Innovation : 14th IFIP WG 6.11 Conference on e-Business, e-Services, and e-Society, I3E 2015. 2015.

[Sh14] Shostack, Adam: Threat modeling: Designing for security. 2014.

[SI] SIOC project website. https://sioc.eu/.

[SKR12] Sabouri, Ahmad; Krontiris, Ioannis; Rannenberg, Kai: Attribute-Based Credentials for Trust (ABC4Trust). In: TrustBus. 2012.

[St16] Statista: , E-Commerce in Deutschland. https://de.statista.com/outlook/243/137/ecommerce/deutschland/, 2016.

[WJ15] Wuyts, Kim; Joosen, Wouter: LINDDUN privacy threat modeling: a tutorial. 2015.

All websites have been last accessed on Dec. 11th, 2017.

C.4 JonDonym Users' Information Privacy Concerns

David Harborth and Sebastian Pape. JonDonym users' information privacy concerns. In *ICT Systems Security and Privacy Protection - 33rd IFIP TC 11 International Conference, SEC 2018, Held at the 24th IFIP World Computer Congress, WCC 2018, Poznan, Poland, September 18-20, 2018, Proceedings*, pages 170–184, 2018. doi: 10.1007/978-3-319-99828-2_13. URL https://doi.org/10.1007/978-3-319-99828-2_13

JonDonym Users' Information Privacy Concerns

David Harborth and Sebastian Pape(✉)

Chair of Mobile Business and Multilateral Security,
Goethe University, Frankfurt, Germany
sebastian.pape@m-chair.de

Abstract. Privacy concerns as well as trust and risk beliefs are important factors that can influence users' decision to use a service. One popular model that integrates these factors is relating the Internet Users Information Privacy Concerns (IUIPC) construct to trust and risk beliefs. However, studies haven't yet applied it to a privacy enhancing technology (PET) such as an anonymization service. Therefore, we conducted a survey among 416 users of the anonymization service JonDonym [1] and collected 141 complete questionnaires. We rely on the IUIPC construct and the related trust-risk model and show that it needs to be adapted for the case of PETs. In addition, we extend the original causal model by including trust beliefs in the anonymization service provider and show that they have a significant effect on the actual use behavior of the PET.

Keywords: Internet Users' Information Privacy Concerns · IUIPC
Anonymity services · Privacy concerns · Trust beliefs · Risk beliefs

1 Introduction

Privacy concerns have been discussed since the very beginning of computer sharing [2]. With a raising economic interest in the internet [3], they gain importance. Bruce Schneier [4] states: "Surveillance is the business model of the internet. Everyone is under constant surveillance by many companies, ranging from social networks like Facebook to cellphone providers." Thus, it can not be a surprise that users have privacy concerns and feel a strong need to protect their privacy[1].

One popular model for measuring and explaining privacy concerns of online users is the Internet Users' Information Privacy Concerns (IUIPC) construct by Malhotra et al. [6]. Their research involves a theoretical framework and an instrument for operationalizing privacy concerns, as well as a causal model for this construct including trust and risk beliefs about the online companies' data handling of personal information. The IUIPC construct has been used in various

[1] "The mean value for the statement'I feel very strongly about protecting my privacy' was 3.64 on a five-point scale with no statistically significant differences across gender, income groups, educational levels, or political affiliation" [5].

L. J. Janczewski and M. Kutyłowski (Eds.): SEC 2018, IFIP AICT 529, pp. 170–184, 2018.
https://doi.org/10.1007/978-3-319-99828-2_13

contexts, e.g. Internet of Things [7], internet transactions [8] and Mobile Apps [9], but to the best of our knowledge the IUIPC construct has never been applied to a privacy enhancing technology (PET) such as anonymization services. The IUIPC instrument shows its strengths best when a service with a certain use for the customer (primary use) is investigated with respect to privacy concerns. However, for anonymization services the primary purpose is to help users to protect their privacy. As a consequence, it is necessary to distinguish between trust and risk beliefs with respect to technologies which aim to protect personal (PETs) and regular internet services. Therefore, the trust model within IUIPC's causal model needs to be adapted for the investigation of anonymization services. For that purpose, we conducted a survey among 416 users of the anonymization service JonDonym [1] and collected 141 complete questionnaires. Our results contribute to the understanding of users' perceptions about PETs and indicate how privacy concerns and trust and risk beliefs influence the use behavior of PETs.

The remainder of the paper is structured as follows: Sect. 2 briefly introduces the JonDonym anonymization service and lists related work on PETs. In Sect. 3, we present the research hypotheses and describe the questionnaire and the data collection process. We assess the quality of our empirical results with regard to reliability and validity in Sect. 4. In Sect. 5, we discuss the implications of the results, elaborate on limitations of the framework and conclude the paper with suggestions for future work.

2 Background and Related Work

Privacy-Enhancing Technologies (PETs) is an umbrella term for different privacy protecting technologies. Borking and Raab define PETs as a "coherent system of ICT measures that protects privacy [...] by eliminating or reducing personal data or by preventing unnecessary and/or undesired processing of personal data; all without losing the functionality of the data system" [10, S. 1].

In this paper, we investigate the privacy, trust and risk beliefs associated with PETs for the case of the anonymity service JonDonym [1]. Comparable to Tor, JonDonym is an anonymity service. However, unlike Tor, it is a proxy system based on mix cascades. It is available for free with several limitations, like the maximum download speed. In addition, there are different premium rates without these limitations that differ with regard to duration and included data volume. Thus, JonDonym offers several different tariffs and is not based on donations like Tor. The actual number of users is not predictable since the service does not keep track of this. JonDonym is also the focus of an earlier user study on user characteristics of privacy services [11]. However, the focus of the study is rather descriptive and does not focus on users' beliefs and concerns.

Previous non-technical work on PETs considers mainly usability studies and does not primarily focus on privacy concerns and related trust and risk beliefs of PET users. For example, Lee et al. [12] assess the usability of the Tor Launcher and propose recommendations to overcome the found usability issues.

Benenson et al. [13] investigate acceptance factors for anonymous credentials. Among other things, they find that trust in the PET has no statistically significant impact on the intention to use the service. This result is relevant for our study since we also hypothesize that trust in JonDonym has a positive effect on the actual use of the service (see Sect. 3.1). Janic et al. [14] claim to consider the relationship between privacy concerns, transparency enhancing technologies (TETs) and PETs, but have a strong focus on TETs and only provide a literature review.

3 Methodology

We base our research on the Internet Users' Information Privacy Concerns (IUIPC) model by Malhotra et al. [6]. The original research on this model investigates the role of users' information privacy concerns in the context of releasing personal information to a marketing service provider. Since we want to investigate the role of privacy concerns, trust and risk beliefs for using a PET (i.e. JonDonym), we can adapt the model by substituting the behavioral intention to perform an action with the actual use of JonDonym. This is possible since we asked current users of JonDonym who actively use the PET. In addition, we extend the original model by trusting beliefs in the PET itself. We argue that the level of trust in a PET is a crucial factor determining the use decision.

For analyzing the cause-effect relationships between the latent (unobserved) variables, we use structural equation modelling (SEM). There are two main approaches for SEM, namely covariance-based SEM (CB-SEM) and partial least squares SEM (PLS-SEM) [15]. Since our research goal is to predict the target construct actual use behavior of JonDonym, we use PLS-SEM for our analysis [15,16]. In the following subsections, we discuss the hypotheses based on the IUIPC model [6], the questionnaire and the data collection process.

3.1 Research Hypotheses

As Fig. 1 shows, the structural model contains several relationships between exogenous and endogenous variables. We develop our research hypotheses for these relationships based on the original hypotheses of the IUIPC model [6]. In the original article, IUIPC is operationalized as a second-order construct of the sub-constructs collection (COLL), awareness (AWA) and control (CONTROL)[2]. Thus, the privacy concerns of users are determined by their concerns about "[...] individual-specific data possessed by others relative to the value of benefits receive" [6, p. 338], the control they have over their own data (i.e. possibilities to change or opt-out) and the "[...] degree to which a consumer is concerned about his/her awareness of organizational information privacy practices" [6, p. 339].

The effect of IUIPC on the behavioral intention (in our model the actual use behavior) is moderated by trusting beliefs and risk beliefs. Trusting beliefs

[2] Due to space limitations, we will not elaborate on the statistics of second-order constructs here. For an extensive discussion see Steward and Malhotra [6,17].

represent users' perceptions about the behavior of online firms to protect the users' personal information. In contrast, risk beliefs represent users' perception about losses associated with providing personal data to online firms [6]. Thus, the higher the privacy concerns of a user, the lower are his or her trusting beliefs and the higher are his or her risk beliefs. In addition, a higher level of trust is assumed to decrease the risk beliefs. Thus, we derive the following three hypotheses:

H 1: *Internet Users' Information Privacy Concerns (IUIPC) have a negative effect on Trusting Beliefs (TB).*
H 2: *Internet Users' Information Privacy Concerns (IUIPC) have a positive effect on Risk Beliefs (RB).*
H 3: *Trusting Beliefs (TB) have a negative effect on Risk Beliefs (RB).*

Since we investigate the use of a specific PET, JonDonym, we extend the model by including the trust of users in JonDonym itself. For that purpose, we adapt the trust construct by Pavlou [18]. However, in order to protect their privacy, users with higher privacy concerns are assumed to rather trust the privacy-enhancing technology compared to online firms that process personal data. In particular, because we surveyed users of the PET. Therefore, we hypothesize:

H 4: *Internet Users Information Privacy Concerns (IUIPC) have a positive effect on the trusting beliefs in JonDonym (TB_{JD}).*

Trust is an important factor in the acceptance decision of users [18]. Especially for the case of privacy protection, we assume that trust in JonDonym is a major factor in the decision to use the technology. Thus, we hypothesize that:

H 5: *Trusting beliefs in JonDonym (TB_{JD}) have a positive effect on the actual use behavior of JonDonym (USE).*

When considering the effects of trusting and risk beliefs on behavior in the context of releasing data to online companies, it is logical that trusting beliefs have a positive effect and risk beliefs have a negative effect on releasing data. However, in our case with actual use behavior of a PET, we assume these effects reverse. The higher the trusting beliefs in online firms, the lower is the use frequency of JonDonym, since the protection of data becomes less important. Following this rationale, a higher degree of risk beliefs with respect to the data processing of online firms leads to a higher degree of use. Therefore, we hypothesize that:

H 6: *Trusting beliefs (TB) have a negative effect on actual use behavior of JonDonym (USE).*

H 7: *Risk beliefs (RB) have a positive effect on actual use behavior of JonDonym (USE).*

3.2 Questionnaire Composition and Data Collection Procedure

The questionnaire constructs are adapted from the original IUIPC paper [6]. We conducted the study with German and English speaking JonDonym users. Thus, we administered two questionnaires. All items for the German questionnaire had to be translated into German since all of the constructs are adapted from English literature. To ensure content validity of the translation, we followed a rigorous translation process [19,20]. First, we translated the English questionnaire into German with the help of a certified translator (translators are standardized following the DIN EN 15038 norm). The German version was then given to a second independent certified translator who retranslated the questionnaire to English. This step was done to ensure the equivalence of the translation. Third, a group of five academic colleagues checked the two English versions with regard to this equivalence. All items were found to be equivalent. The items of the English version can be found in Appendix A.

Since we investigate the effect of privacy concerns, trust and risk beliefs on the use of JonDonym, we collected data of actual users of the PET. We installed the surveys on a university server and managed it with the survey software LimeSurvey (version 2.63.1) [21]. The links to the English and German version were distributed with the beta version of the JonDonym browser and published on the official JonDonym homepage. In sum, 416 participants started the questionnaire (173 for the English version and 243 for the German version). Of those 416 approached participants, 141 (53 for the English version and 88 for the German version) remained after deleting unfinished sets and all participants who answered a test question in the middle of the survey incorrectly.

The demographic questions were not mandatory to fill out. This was done on purpose since we assumed that most of the participants are highly sensitive with respect to their personal data. Therefore, we resign from a discussion of the demographics in our research context. This decision is backed up by Singh and Hill, who found no statistically significant differences across gender, income groups, educational levels, or political affiliation in the desire to protect one's privacy [5].

4 Results

We tested the model using SmartPLS version 3.2.6 [22]. Before looking at the result of the structural model and discussing its implications, we discuss the measurement model, and check for the reliability and validity of our results. This is a precondition of being able to interpret the results of the structural model. Furthermore, it is recommended to report the computational settings. For the PLS algorithm, we choose the path weighting scheme with a maximum of 300 iterations and a stop criterion of 10^{-7}. For the bootstrapping procedure, we use 5000 bootstrap subsamples and no sign changes as the method for handling sign changes during the iterations of the bootstrapping procedure.

4.1 Assessment of the Measurement Model

As the model is measured solely reflectively, we need to evaluate the internal consistency reliability, convergent validity and discriminant validity to assess the measurement model properly [15].

Internal Consistency Reliability. Internal consistency reliability (ICR) measurements indicate how well certain indicators of a construct measure the same latent phenomenon. Two standard approaches for assessing ICR are Cronbach's α and the composite reliability. The values of both measures should be between 0.7 and 0.95 for research that builds upon accepted models. Values of Cronbach's α are seen as a lower bound and values of the composite reliability as an upper bound of the assessment [16]. Table 1 includes the ICR of the variables in the last two rows. It can be seen that all values for Cronbach's α are above the lower threshold of 0.7 except for RB. However, for the composite reliability the value for RB is higher than 0.7. Therefore, we argue that ICR is not an issue for this variable. For all variables, no value is above 0.95. Values above that upper threshold indicate that the indicators measure the same dimension of the latent variable, which is not optimal with regard to the validity [16]. In sum, ICR is established for our variables. The variables IUIPC and USE are single-item constructs, and thus have ICR values of 1.

Convergent Validity. Convergent validity determines the degree to which indicators of a certain reflective construct are explained by that construct. This is assessed by calculating the outer loadings of the indicators of the constructs (indicator reliability) and by looking at the average variance extracted (AVE) [15]. Loadings above 0.7 imply that the indicators have much in common, which is desirable for reflective measurement models [16]. Table 1 shows the outer loadings in bold on the diagonal. All loadings are higher than 0.7, except for RISK5 and TB5. Since the AVE of these constructs is still above 0.5, we do not drop these items. Convergent validity for the construct is assessed by the AVE. AVE is equal to the sum of the squared loadings divided by the number of indicators. A threshold of 0.5 is acceptable, indicating that the construct explains at least half of the variance of the indicators [16]. The diagonal values of Table 2 present the AVE of our constructs. All values are well above 0.5, demonstrating convergent validity.

Discriminant Validity. Discriminant validity measures the degree of uniqueness of a construct compared to other constructs. Comparable to the convergent validity assessment, two approaches are used for investigated discriminant validity. The first approach, assessing cross-loadings, is dealing with single indicators. All outer loadings of a certain construct should be larger than its cross-loadings with other constructs [15]. Table 1 illustrates the cross-loadings as off-diagonal elements. All cross-loadings are smaller than the outer loadings, fulfilling the first assessment approach of discriminant validity. The second approach is on the construct level and compares the square root of the constructs' AVE with the correlations with other constructs. The square root of the AVE of a single

Table 1. Loadings and cross-loadings of the reflective items and internal consistency reliability

Constructs	AWA	CONTROL	COLL	RB	TB	TB_{JD}	IUIPC	USE
AWA1	**0.892**	0.254	0.297	0.050	−0.107	0.073	0.614	0.143
AWA2	**0.927**	0.254	0.287	0.072	−0.152	0.057	0.622	0.098
AWA3	**0.883**	0.297	0.356	0.235	−0.207	0.071	0.648	0.169
CONTROL1	0.284	**0.837**	0.379	0.271	−0.306	0.163	0.618	0.208
CONTROL2	0.244	**0.808**	0.238	0.205	−0.075	0.103	0.505	0.175
CONTROL3	0.201	**0.819**	0.348	0.287	−0.195	0.089	0.514	0.138
COLL1	0.202	0.309	**0.781**	0.237	−0.084	0.152	0.588	0.133
COLL2	0.199	0.185	**0.760**	0.141	0.001	0.262	0.548	0.300
COLL3	0.380	0.364	**0.873**	0.192	−0.063	0.297	0.733	0.302
COLL4	0.336	0.416	**0.872**	0.349	−0.213	0.193	0.720	0.261
RB1	0.117	0.213	0.230	**0.814**	−0.324	0.022	0.194	0.157
RB2	0.061	0.172	0.100	**0.710**	−0.201	−0.114	0.116	0.050
RB3	0.132	0.225	0.193	**0.815**	−0.179	−0.098	0.196	0.123
RB4	0.075	0.214	0.266	**0.811**	−0.241	−0.076	0.211	0.050
RB5	−0.112	−0.311	−0.244	**−0.682**	0.392	0.050	−0.277	−0.092
TB1	−0.174	−0.217	−0.078	−0.296	**0.832**	0.028	−0.196	−0.117
TB2	−0.114	−0.171	−0.033	−0.281	**0.835**	−0.101	−0.130	−0.134
TB3	−0.167	−0.210	−0.116	−0.343	**0.815**	0.004	−0.209	−0.024
TB4	−0.123	−0.160	−0.089	−0.212	**0.666**	−0.051	−0.129	−0.060
TB5	−0.121	−0.210	−0.137	−0.354	**0.855**	−0.158	−0.200	−0.210
$TB_{JD}1$	0.017	0.104	0.244	−0.058	−0.100	**0.898**	0.130	0.281
$TB_{JD}2$	0.088	0.117	0.222	−0.109	−0.043	**0.922**	0.165	0.303
$TB_{JD}3$	0.090	0.176	0.284	−0.032	−0.060	**0.922**	0.199	0.330
IUIPC	0.698	0.669	0.794	0.276	−0.220	0.183	**1.000**	0.333
USE	0.152	0.214	0.304	0.130	−0.142	0.335	0.333	**1.000**
Cronbach's α	0.883	0.761	0.841	0.612	0.862	0.902	1.000	1.000
Composite reliability	0.928	0.862	0.893	0.749	0.901	0.938	1.000	1.000

construct should be larger than the correlation with other constructs (Fornell-Larcker criterion) [16]. Table 2 contains the square root of the AVE on the diagonal in parentheses. All values are larger than the correlations with other constructs, indicating discriminant validity. Since there are problems in determining the discriminant validity with both approaches, researchers propose the heterotrait-monotrait ratio (HTMT) for assessing discriminant validity as a superior approach to the former ones [23]. HTMT divides between-trait correlations by within-trait correlations, therefore providing a measure of what the true correlation of two constructs would be if the measurement is flawless [16]. Values close to 1 for HTMT indicate a lack of discriminant validity. A conservative

Table 2. Discriminant validity with AVEs and construct correlations

Constructs (AVE)	AWA	COLL	CONTROL	IUIPC	RB	TB	TB_{JD}	USE
AWA (0.811)	0.901							
COLL (0.678)	0.349	0.823						
CONTROL (0.675)	0.298	0.396	0.822					
IUIPC (1.000)	0.698	0.794	0.669	1,000				
RB (0.591)	0.134	0.284	0.311	0.276	0.769			
TB (0.646)	−0.173	−0.116	−0.243	−0.220	−0.377	0.804		
TB_{JD} (0.835)	0.074	0.275	0.148	0.183	−0.071	−0.072	0.914	
USE (1.000)	0.152	0.304	0.214	0.333	0.130	−0.142	0.335	1.000

Note: AVEs in parentheses in the first column. Values for $\sqrt{AVE}$ are shown on the diagonal and construct correlations are off-diagonal elements.

threshold is 0.85 [23]. Table 3 contains the values for HTMT and no value is above the suggested threshold of 0.85.

Table 3. Heterotrait-monotrait ratio (HTMT)

Constructs	AWA	COLL	CONTROL	IUIPC	RB	TB	TB_{JD}	USE
AWA								
COLL	0.393							
CONTROL	0.360	0.478						
IUIPC	0.742	0.858	0.761					
RB	0.155	0.313	0.368	0.282				
TB	0.198	0.142	0.287	0.232	0.402			
TB_{JD}	0.091	0.314	0.171	0.190	0.109	0.118		
USE	0.161	0.330	0.242	0.333	0.133	0.146	0.351	

To evaluate whether the HTMT statistics are significantly different from 1, a bootstrapping procedure with 5,000 subsamples is conducted to get the confidence interval in which the true HTMT value lies with a 95% chance. The HTMT measure requires that no confidence interval contains the value 1. The conducted analysis shows that this is the case. Thus, discriminant validity is established for our model.

Common Method Bias. The common method bias (CMB) can occur if data is gathered with a self-reported survey at one point in time in one questionnaire [24]. Since this is the case in our research design, the need to test for CMB arises.

An unrotated principal component factor analysis is performed with the software package STATA 14.0 to conduct the Harman's single-factor test to address the issue of CMB [25]. The assumptions of the test are that CMB is not an issue if there is no single factor that results from the factor analysis or that the first

factor does not account for the majority of the total variance [25]. The test shows that six factors have eigenvalues larger than 1 which account for 69.45% of the total variance. The first factor explains 23.74% of the total variance. Based on the results of previous literature [26], we argue that CMB is not likely to be an issue in the data set.

4.2 Assessment and Results of the Structural Model

To assess the structural model, we follow the steps proposed by Hair et al. [16] which include an assessment of possible collinearity problems, of path coefficients, of the level of R^2, of the effect size f^2, of the predictive relevance Q^2 and the effect size q^2. We address these evaluation steps to ensure the predictive power of the model with regard to the target constructs.

Collinearity. Collinearity is present if two predictor variables are highly correlated with each other. To address this issue, we assess the inner variance inflation factor (inner VIF). All VIF values above 5 indicate that collinearity between constructs is present. For our model, the highest VIF is 1.179. Thus collinearity is apparently not an issue.

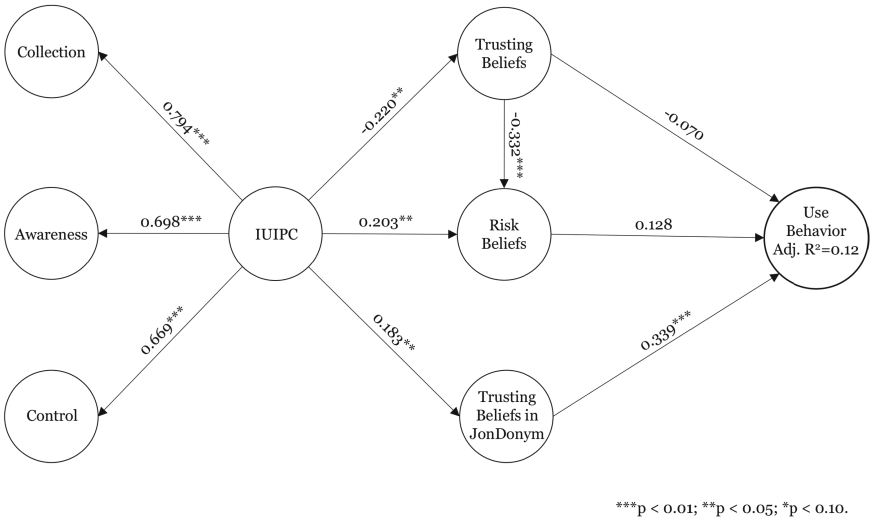


Fig. 1. Path estimates and adjusted R^2 values of the structural model

Significance and Relevance of Model Relationships. Figure 1 presents the results of the path estimations and the adjusted R^2 of the endogenous variable USE. We used the adjusted R^2 as it is a conservative measure for the explained variance of a dependent variable by avoiding a bias towards more complex models [16]. The R^2 is 0.12 for USE. Thus, our models explains 12% of the variance in USE.

There are different proposals for interpreting the size of this value. We choose to use the very conservative threshold proposed by Hair et al. [15], where R^2 values are weak with values around 0.25, moderate with 0.50 and substantial with 0.75. Based on this classification, the R^2 value for USE is rather weak. The path coefficients are presented on the arrows connecting the exogenous and endogenous constructs in Fig. 1. Statistical significance is indicated by asterisks, ranging from three asterisks for p-values smaller than 0.01 to one asterisk for p-values smaller than 0.10. The p-value indicates the probability that a path estimate is incorrectly assumed to be significant. Thus, the lower the p-value, the higher the probability that the given relationship exists. The relevance of the path coefficients is expressed by the relative size of the coefficient compared to the other explanatory variables [16].

It can be seen that IUIPC has a statistically significant negative medium-sized effect on trusting beliefs and a positive effect on risk beliefs. The effect of IUIPC on trusting beliefs in JonDonym is significant, positive and medium-sized. The construct trusting beliefs has a statistically significant medium-sized negative effect on risk beliefs. The effect of trusting beliefs on use behavior is negative, but not statistically significant. The same holds for the relationship between risk beliefs and use behavior (for both $p \geq 0.10$). In contrast, the effect of trusting beliefs in JonDonym on use behavior is highly statistically significant, positive and large with 0.339.

Effect Sizes f^2. The f^2 effect size measures the impact of a construct on the endogenous variable by omitting it from the analysis and assessing the resulting change in the R^2 value [16]. The values are assessed based on thresholds by Cohen [27], who defines effects as small, medium and large for values of 0.02, 0.15 and 0.35, respectively. Table 4 shows the results of the f^2 evaluation. Values in italics indicate small effects and values in bold indicate medium effects. All other values have no substantial effect. The results correspond to those of the previous analysis of the path coefficients.

Table 4. Values for the f^2 and q^2 effect size assessment

Variables	f^2	q^2
Endogenous / Exogenous	USE	USE
RB	0.016	0.012
TB	0.005	-0.016
TB_{JD}	*0.131*	*0.109*

Predictive Relevance Q^2. The Q^2 measure indicates the out-of-sample predictive relevance of the structural model with regard to the endogenous latent variables based on a blindfolding procedure [16]. We used an omission distance $d = 7$. Recommended values for d are between five and ten [15]. Furthermore, we report the Q^2 values of the cross-validated redundancy approach, since this approach is

based on both the results of the measurement model as well as of the structural model [16]. Detailed information about the calculation cannot be provided due to space limitations. For further information see Chin [28]. For our model, Q^2 is calculated for USE. Values above 0 indicate that the model has the property of predictive relevance. In our case, the Q^2 value is equal to 0.097 for USE. Since they are larger than 0, predictive relevance of the model is established.

Effect Sizes q^2. The assessment of q^2 follows the same logic as the one of f^2. It is based on the Q^2 values of the endogenous variables and calculates the individual predictive power of the exogenous variables by omitting them and comparing the change in Q^2. The effect sizes q^2 have to be calculated with the formula [16]:

$$q^2_{X \to Y} = \frac{Q^2_{included} - Q^2_{excluded}}{1 - Q^2_{included}}$$

All individual values for q^2 are calculated with an omission distance d of seven. The results are shown in Table 4. The thresholds for the f^2 interpretation can be applied here, too [27]. Values in italics indicate small effects and values in bold indicate medium effects. All other values have no substantial effect. As before, only the trust in JonDonym has a medium-sized effect, implying the highest predictive power of all included exogenous variables.

5 Discussion and Conclusion

Based on our results, hypotheses H1 to H5 can be confirmed, whereas H6 and H7 cannot be confirmed (cf. Table 5). The results for H6 and H7 are very surprising, considering that they are in contrast to the rationale explained in Sect. 3.1 and the results from previous literature [6]. However, it must be said that it is possible that the relatively small sample size of 141 leads to a statistical non-significance when effect sizes are rather small. Therefore, we cannot rule out that the effects of risk beliefs and trusting beliefs on use would be significant with a larger sample size. Thus, only the degree of trust in the PET (JonDonym) has a significant and large effect on the use behavior. This result shows that it is crucial for a PET provider to establish a trustful reputation to get used. The trusting beliefs in the PET itself are positively influenced by the users' information privacy concerns. Thus, the results imply that users with a higher level of privacy concerns rather tend to trust a PET. The limitations of the study primarily concern the sample composition and size. First, a larger sample would have been beneficial. However, in general, a sample of 141 participants is acceptable for our kind of statistical analysis [16] and active users of a PET are hard to find for a relatively long online questionnaire. This is especially the case, if they do not have any financial rewards as in our study. Second, the combination of the results of the German and the English questionnaire can be a potential source for errors. Participants might have understood the questionnaire in German differently than the participants who filled out the English version. We argue that we achieved equivalence with regard to the meaning through conducting a thorough translation process, and therefore limiting this potential source of error to the largest

extent possible. In addition, combining the data was necessary from a pragmatic point of view to get a sample size as large as possible for the statistical analysis.

Further work is required to investigate the specific determinants of use decisions for or against PETs and break down the interrelationships between the associated antecedents. In particular, it would be interesting to investigate the relationship between trusting beliefs in online companies and trust in the PET itself. A theoretical underlying is required to include this relationship in our structural equation model.

In this paper, we contributed to the literature on privacy-enhancing technologies and users' privacy by assessing the specific relationships between information privacy concerns, trusting beliefs in online firms and a privacy-enhancing technology (in our case JonDonym), risk beliefs associated with online firms data processing and the actual use behavior of JonDonym. By adapting and extending the IUIPC model by Malhotra et al. [6], we could show that several of the assumptions for regular online services do not hold for PETs.

Table 5. Summary of the results

Hypothesis	Result
H1: Internet Users Information Privacy Concerns (IUIPC) have a negative effect on Trusting Beliefs (TB)	✓
H2: Internet Users Information Privacy Concerns (IUIPC) have a positive effect on Risk Beliefs (RB)	✓
H3: Trusting Beliefs (TB) have a negative effect on Risk Beliefs (RB)	✓
H4: Internet Users Information Privacy Concerns (IUIPC) have a positive effect on the trusting beliefs in JonDonym (TB_{JD})	✓
H5: Trusting beliefs in JonDonym (TB_{JD}) have a positive effect on the actualuse behavior of JonDonym (USE)	✓
H6: Trusting beliefs (TB) have a negative effect on actual use behavior of JonDonym (USE)	✗
H7: Risk beliefs (RB) have a positive effect on actual use behavior of JonDonym (USE)	✗

Acknowledgments. This research was partly funded by the German Federal Ministry of Education and Research (BMBF) with grant number: 16KIS0371. In addition, we thank Rolf Wendolski (JonDos GmbH) for his help during the data collection process.

A Questionnaire

The following items are measured with a seven-point Likert scale, ranging from "strongly disagree" to "strongly agree".

Collection (COLL)

1. It usually bothers me when online companies ask me for personal information.
2. When online companies ask me for personal information, I sometimes think twice before providing it.
3. It bothers me to give personal information to so many online companies.
4. I'm concerned that online companies are collecting too much personal information about me.

Awareness (AWA)

1. Companies seeking information online should disclose the way the data are collected, processed, and used.
2. A good consumer online privacy policy should have a clear and conspicuous disclosure.
3. It is very important to me that I am aware and knowledgeable about how my personal information will be used.

Control (CONTROL)

1. Consumer online privacy is really a matter of consumers? right to exercise control and autonomy over decisions about how their information is collected, used, and shared.
2. Consumer control of personal information lies at the heart of consumer privacy.
3. I believe that online privacy is invaded when control is lost or unwillingly reduced as a result of a marketing transaction.

Trusting Beliefs (TB)

1. Online companies are trustworthy in handling information.
2. Online companies tell the truth and fulfill promises related to information provided by me.
3. I trust that online companies would keep my best interests in mind when dealing with information.
4. Online companies are in general predictable and consistent regarding the usage of information.
5. Online companies are always honest with customers when it comes to using the provided information.

Risk Beliefs (RB)

1. In general, it would be risky to give information to online companies.
2. There would be high potential for loss associated with giving information to online firms.
3. There would be too much uncertainty associated with giving information to online firms.
4. Providing online firms with information would involve many unexpected problems.
5. I would feel safe giving information to online companies.

Trusting Beliefs in JonDonym ($\mathbf{TB_{JD}}$)

1. JonDonym ist trustworthy.
2. JonDonym keeps promises and commitments.
3. I trust JonDonym because they keep my best interests in mind.

Use Behavior (USE)

1. Please choose your usage frequency for JonDonym[3]
   - Never
   - Once a month
   - Several times a month
   - Once a week
   - Several times a week
   - Once a day
   - Several times a day
   - Once an hour
   - Several times an hour
   - All the time.

[3] The frequency scale is adapted from Rosen et al. [29].

References

1. JonDos Gmbh: Official Homepage of JonDonym (2018). https://www.anonym-surfen.de
2. David, E.E., Fano, R.M.: Some thoughts about the social implications of accessible computing. In: Proceedings 1965 Fall Joint Computer Conference (1965). http://www.multicians.org/fjcc6.html
3. Bédard, M.: The Underestimated Economic Benefits of the Internet. Economic Notes, Regulation series. The Montreal Economic Institute, Montreal (2016)
4. Mineo, L.: On internet privacy, be very afraid (Interview with Bruce Schneier), August 2017. https://news.harvard.edu/gazette/story/2017/08/when-it-comes-to-internet-privacybe-very-afraid-analyst-suggests/
5. Singh, T., Hill, M.E.: Consumer privacy and the internet in Europe: a view from Germany. J. Consum. Mark. **20**(7), 634–651 (2003)
6. Malhotra, N.K., Kim, S.S., Agarwal, J.: Internet users' information privacy concerns (IUIPC): the construct, the scale, and a causal model. Inf. Syst. Res. **15**(4), 336–355 (2004)
7. Naeini, P.E., et al.: Privacy expectations and preferences in an IoT world. In: Symposium on Usable Privacy and Security (SOUPS) (2017)
8. Heales, J., Cockcroft, S., Trieu, V.-H.: The Influence of privacy, trust, and national culture on internet transactions. In: Meiselwitz, G. (ed.) SCSM 2017. LNCS, vol. 10282, pp. 159–176. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-58559-8_14
9. Raber, F., Krueger, A.: Towards understanding the influence of personality on mobile app permission settings. In: Bernhaupt, R., Dalvi, G., Joshi, A., K. Balkrishan, D., ONeill, J., Winckler, M. (eds.) INTERACT 2017. LNCS, vol. 10516, pp. 62–82. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-68059-0_4
10. Borking, J.J., Raab, C.: Laws, PETs and other technologies for privacy protection. J. Inf. Law Technol. **1**, 1–14 (2001)
11. Spiekermann, S.: The desire for privacy: insights into the views and nature of the early adopters of privacy services. Int. J. Technol. Hum. Interact. **1**(1), 74–83 (2005)
12. Lee, L., Fifield, D., Malkin, N., Iyer, G., Egelman, S., Wagner, D.: A usability evaluation of tor launcher. In: Proceedings on Privacy Enhancing Technologies, no. 3, pp. 90–109 (2017)
13. Benenson, Z., Girard, A., Krontiris, I.: User acceptance factors for anonymous credentials: an empirical investigation. In: 14th Annual Workshop on the Economics of Information Security (WEIS), pp. 1–33 (2015)
14. Janic, M., Wijbenga, J.P., Veugen, T.: Transparency enhancing tools (tets): an overview. In: 2013 Third Workshop on Socio-Technical Aspects in Security and Trust (STAST), pp. 18–25. IEEE (2013)
15. Hair, J., Ringle, C.M., Sarstedt, M.: PLS-SEM: indeed a silver bullet. J. Mark. Theory Pract. **19**(2), 139–152 (2011)
16. Hair, J., Hult, G.T.M., Ringle, C.M., Sarstedt, M.: A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM). SAGE Publications, Thousand Oaks (2017)
17. Stewart, K.A., Segars, A.H.: An empirical examination of the concern for information privacy instrument. Inf. Syst. Res. **13**(1), 36–49 (2002)
18. Pavlou, P.A.: Consumer acceptance of electronic commerce: integrating trust and risk with the technology acceptance model. Int. J. Electron. Commer. **7**(3), 101–134 (2003)

19. Harborth, D., Pape, S.: Exploring the hype: investigating technology acceptance factors of Pokémon Go. In: 2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), pp. 155–168 (2017)
20. Harborth, D., Pape, S.: Privacy concerns and behavior of Pokémon go players in Germany. In: Hansen, M., Kosta, E., Nai-Fovino, I., Fischer-Hübner, S. (eds.) Proceedings of IFIP Summer School on Privacy and Identity Management (IFIPSC2017), pp. 314–329. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-92925-5_21
21. Schmitz, C.: LimeSurvey Project Team (2015). http://www.limesurvey.org
22. Ringle, C.M., Wende, S., Becker, J.M.: SmartPLS 3 (2015). http://www.smartpls.com
23. Henseler, J., Ringle, C.M., Sarstedt, M.: A new criterion for assessing discriminant validity in variance-based structural equation modeling. J. Acad. Mark. Sci. **43**(1), 115–135 (2015)
24. Malhotra, N.K., Kim, S.S., Patil, A.: Common method variance in IS research: a comparison of alternative approaches and a reanalysis of past research. Manag. Sci. **52**(12), 1865–1883 (2006)
25. Podsakoff, P.M., MacKenzie, S.B., Lee, J.Y., Podsakoff, N.P.: Common method biases in behavioral research: a critical review of the literature and recommended remedies. J. Appl. Psychol. **88**(5), 879–903 (2003)
26. Blome, C., Paulraj, A.: Ethical climate and purchasing social responsibility: a benevolence focus. J. Bus. Eth. **116**(3), 567–585 (2013)
27. Cohen, J.: Statistical Power Analysis for the Behavioral Sciences. Lawrence Earlbaum Associates, HillsDale (1988)
28. Chin, W.W.: The partial least squares approach to structural equation modeling. In: Marcoulides, G.A. (ed.): Modern Methods for Business Research, pp. 295–336. Lawrence Erlbaum, Mahwah (1998)
29. Rosen, L., Whaling, K., Carrier, L., Cheever, N., Rokkum, J.: The media and technology usage and attitudes scale: an empirical investigation. Comput. Hum. Behav. **29**(6), 2501–2511 (2013)

C.5 Assessing Privacy Policies of Internet of Things Services

Niklas Paul, Welderufael B. Tesfay, Dennis-Kenji Kipker, Mattea Stelter, and Sebastian Pape. Assessing privacy policies of internet of things services. In *ICT Systems Security and Privacy Protection - 33rd IFIP TC 11 International Conference, SEC 2018, Held at the 24th IFIP World Computer Congress, WCC 2018, Poznan, Poland, September 18-20, 2018, Proceedings*, pages 156–169, 2018. doi: 10.1007/978-3-319-99828-2_12. URL https://doi.org/10.1007/978-3-319-99828-2_12

Assessing Privacy Policies of Internet of Things Services

Niklas Paul[1], Welderufael B. Tesfay[1], Dennis-Kenji Kipker[2], Mattea Stelter[2], and Sebastian Pape[1(✉)]

[1] Goethe-University, Frankfurt, Germany
sebastian.pape@m-chair.de
[2] University of Bremen, Bremen, Germany

Abstract. This paper provides an assessment framework for privacy policies of Internet of Things Services which is based on particular GDPR requirements. The objective of the framework is to serve as supportive tool for users to take privacy-related informed decisions. For example when buying a new fitness tracker, users could compare different models in respect to privacy friendliness or more particular aspects of the framework such as if data is given to a third party. The framework consists of 16 parameters with one to four yes-or-no-questions each and allows the users to bring in their own weights for the different parameters. We assessed 110 devices which had 94 different policies. Furthermore, we did a legal assessment for the parameters to deal with the case that there is no statement at all regarding a certain parameter. The results of this comparative study show that most of the examined privacy policies of IoT devices/services are insufficient to address particular GDPR requirements and beyond. We also found a correlation between the length of the policy and the privacy transparency score, respectively.

Keywords: Internet of Things · Privacy policies
General Data Protection Regulation · GDPR · ePrivacy Regulation
ePR

1 Introduction

Privacy is a big but early stage research topic in the Internet of Things (IoT), where many questions are still inadequately addressed [1]. Studies indicate that "six in ten Internet of Things devices don't properly tell customers how their personal information is being used" [2] and "nearly all areas (of Internet of Things) miss applicable mechanisms in privacy" [3]. This collection and processing of personal, sometimes sensitive, information has raised privacy concerns of users. A survey in 2016 revealed that 53% of 797 IT professionals are very concerned

L. J. Janczewski and M. Kutyłowski (Eds.): SEC 2018, IFIP AICT 529, pp. 156–169, 2018.
https://doi.org/10.1007/978-3-319-99828-2_12

about privacy in IoT, it already seems relevant in professional circles [4]. With the increasing complexity of products users have to deal with, it is likely that this raises concerns of non-professional users as well.

Thus, regulators require service providers to publish their data processing practices. As such, terms and conditions and privacy policies are used to inform users about the purpose of data collection and processing. However, only a small proportion of users read these documents [5,6], mainly due to the length of the texts, and being written in difficult legal jargon. Therefore, it is widely accepted to confirm a policy without reading it, even if users in general should read them [7]. As a consequence, users are not aware that a large number of policies elude domestic justice, contains user unfriendly parts or suspect purpose of private data use e.g. to collect information and to use it as "a new source of revenue" by selling the information or for advertising purposes [8].

To give a methodological assessment of this problem, in this work, we introduce a framework for privacy policies of Internet of Things (IoT) devices evaluation based on General Data Protection Regulation (GDPR) aspects as assessment criteria. The framework gives an overview of the contents of certain policies and further ranks them based on their scores pertinent to these criteria. The objective of the framework is not to provide binding legal guidance, but to serve as supportive tool for users to take privacy-related informed decisions. For example when buying a new fitness tracker, users could compare different models in respect to privacy friendliness or more particular aspects of the framework such as if data is given to a third party.

The remainder of the paper is structured as follows: Sect. 2 briefly introduces the regulatory background on which our framework is based. After that, in Sect. 3, related work is presented and how this work differs from them. In Sect. 4 we present our research methodology and in Sect. 5, the assessment framework is introduced. In Sect. 6 we present the results of a first assessment and statistical analyses. In Sect. 7, we discuss results and limitations of the framework and suggest future work. We conclude in Sect. 8.

2 Background

Internet of Things (IoT) refers to the networked interconnection of everyday objects, which are often equipped with ubiquitous intelligence [9]. Usually users can extend the control of IoT devices by using an application on their phone, tablet or computer. Since IoT-Services require a certain amount of personal information to determine user behaviour and they process electronic data automatically, they are regulated by the General Data Protection Regulation (GDPR) [10] and the ePrivacy Regulation (ePR) [11]. In this section, we give a brief overview on the GDPR and ePR with a focus how to utilize them as foundation for the privacy policy assessment framework.

2.1 General Data Protection Regulation

The General Data Protection Regulation, adopted by the European Parliament on 14 April 2016 and becoming effective as from 25 May 2018, will replace the Data Protection Directive (1995/46/EC). The regulation is the result of the EU's objective to harmonize the several data protection provisions existing at European and national level and thereby to strengthen data protection throughout the EU[1]. Unlike the previous directive, the new regulation does not require transposition into national laws and will be directly applicable in all Member States. Henceforth, national legislation that diverges from the GDPR provisions will be allowed only within various opening clauses contained in the regulation. Since the GDPR "lays down rules relating to the protection of natural persons with regard to the processing of personal data" [10, Article 1 para. 1], it is also addressed to suppliers of IoT products. According to Article 3 of the regulation, the GDPR thereby does not only apply for EU-based producers of IoT devices, but also for all enterprises established outside the EU that offer their products on the European market. Therefore, the provisions of the GDPR can serve as uniform assessment criteria for the comparison of the level of data protection ensured for IoT devices whose producers are located across the world.

Of particular importance for the evaluation of privacy policies is Article 13 GDPR, which specifies the information to be provided where personal data are collected from a data subject. These information obligations follow from the transparency principle laid down in Article 5 GDPR. The mandatory information includes, inter alia, identity and contact details of the product provider as well as full details on the purposes of the data processing, the storage period, the various rights of the data subject under Articles 12–23 GDPR, or, where applicable, the disclosure of data to a third party and the transfer of data to third countries.

2.2 ePrivacy Regulation

However, the legislative process on the harmonisation of European data protection law is not yet completed. Apart from the GDPR, the ePrivacy Regulation is intended to replace the outdated Privacy and Electronic Communications Directive (2002/58/EC) and to supplement the GDPR as regards the electronic communication sector. Although the ePrivacy Regulation initially had been expected to become effective at the same time as the GDPR on 25 May 2018, it is currently still at the stage of draft [11]. While trilogue negotiations between the Parliament, the Commission and the Council are about to take place, the high level of data protection provided in the proposal is strongly criticised by media and advertising industries[2]. The exact scope of the ePrivacy Regulation and its relation to the GDPR remain controversial, too [13]. Thus, it does not appear to be appropriate to include the current draft regulation into this assessment framework – the discrepancies that have to be resolved prior to the adoption of a final

[1] See, inter alia, Recitals 6, 7, 9, 10 of the GDPR.
[2] See, for example, the campaign by several industry associations [12].

version are too fundamental. However, in the future, legal requirements for IoT devices will be significantly determined not only by the GDPR, but also by the ePrivacy Regulation: Recital 12 of the proposed regulation explicitly states that the scope of the regulation also covers the transmission of machine-to-machine communications, which is the essential characteristic of the Internet of Things. The regulation's entry into force is not expected before 2019 [14].

3 Related Work

Even though information privacy is a concern for users and IoT operators, so far, it seems to be addressed inadequately. However, there are still some promising efforts, which we summarize below. Stankovic [1] proposed a new language for privacy policies in IoT to address emerging problems of privacy. Ziegeldorf et al. stated seven categories of privacy threats in the Internet of Things, introducing four new categories of privacy threats especially in the Internet of Things [15]. The threat of life-cycle transition (changes of control spheres e.g. through selling) is considered in this framework as well.

Smith, Milberg and Burke found five central dimensions of concerns about privacy practices namely, collection of personal information, internal unauthorized secondary use of personal information, external unauthorized secondary use of personal information and finally errors and improper access [16]. All these previously mentioned dimensions should be addressed in a privacy policy and are also, to some extent, part of the requirements for the assessment framework and can be considered as the basis to develop the framework.

Previous studies examined the existence of policies rather than assessing the content [17]. Previous work that took the content into account, mainly dealt with privacy policies of websites, but not of IoT services and respectively, apps to control them [17–19]. For Example some of them used the Fair Information Practices (FIPs) for the content and the Flesch grade level [20] for assessing the readability with the result that the examined policies were difficult to read and required a higher education level. The Flesch Score is based on the average length of a sentence and the average word length within syllables, the higher it is the easier a text is to read. Over time more mathematical approaches which calculated scores were established but also rankings based on a crowdsourcing approach [19]. In 2017, the project "Ranking Digital Rights" evaluated a set of companies based on 35 parameters in three groups namely governance, freedom of expression and privacy [21]. The privacy category was by far the largest, consisting of 18 parameters. It examined a broad variety of characteristics reaching from simple and easy policy access to the supply of information about potential cyber risks. Noteworthy is, they assessed not only one service of the company but a service portfolio. The project "Terms of Service; Didn't read" uses a less mathematical approach [22]. Based on crowdsourcing they present summaries and a rating of terms of 8 services that are assessed by other users on their website. The problem with this and other crowdsourcing solutions is that the scope is highly dependent on participation [23]. To overcome this, the project"Privee"

uses a combination of crowdsourcing and automated classification [23]. Despite most previous work dealing with website privacy policies, there are also works assessing privacy aspects of apps [24].

4 Methodology

This section briefly describes how the framework was designed, how the assessed policies were selected, and how the assessment procedure was.

4.1 Framework Development

The main goal of this work is to create an assessment framework for privacy policies to assess a large variety of IoT devices. Therefore, applicable parameters are needed. The framework is strongly inspired by the GDPR (cf. Sect. 2), but we also considered the categories of privacy threats from Ziegeldorf et al. [15] and the dimensions of concerns about privacy practices from Smith et al. [16] (cf. Sect. 3). For each of the parameters we identified relevant yes-or-no questions. For all categories, we did a legal assessment to check how we should cope with a non existing statement. We explain this in more detail in Sect. 5.1.

We identified two important dimensions for the framework: (i) Content-Dimension (Privacy Score) and (ii) Transparency-Dimension (Transparency Score). They differ in so far that the transparency-dimension rather checks whether the policy makes a statement or not and the content-dimension rather checks what statement the policy makes.

4.2 Policy Selection

To get an overview of the available products on the market, two websites[3] were used. Since many listed devices didn't exist anymore, we searched in web shops (e.g. Amazon) for similar products. As the framework is built on the GDPR and the GDRP applies only to services provided to EU citizens, the product must be available on the European market. Criteria defining what products are available in terms of the GDPR can be found in Recital 23 [10] and were checked by searching the manufacturers website and web shops. We did not assess policies where we couldn't find the IoT device available to the European market.

Another condition was that the policy needed to be available in English language. If no general EU-English policy was available, an English version applicable in Germany was looked for or otherwise the UK one was chosen. Sometimes, e.g. US policies are slightly different from EU-language policies. If there was an US and an EU policy available, the EU one was chosen. If some parts of the policy were applicable to specific countries, the descriptions for Germany or otherwise another EU-country were preferred. If there was no distinction of EU/Non-EU or no declaration of where the policies apply, it was assumed that it is a global policy, which is also permitted in the framework.

[3] http://IoTLineup.com and http://IoTList.co.

To find the policies we searched the website of the manufacturer in the first place and after that we searched for the policy in the Google Playstore and in the last instance we contacted them via E-Mail to send us the according policy.

4.3 Assessment Procedure

The assessment was done manually by reading the policies and applying all parameters to them. The number of words and the Flesch Score were calculated automatically by an Online Tool [25], the remaining questions are yes-or-no questions. To record the results of the assessment, a table-workbook with several sheets was created containing an overview of all policies and one sheet for every assessment. The assessment scorecard is a table with general information (e.g. name, ID, category) in the header and all parameters beneath. For both Privacy Score and Transparency Score there are columns where the answer and the corresponding points were saved. We also stored the segment of the privacy policy which was relevant for the scoring to allow using this data as a training set for a machine learning algorithm later.

5 Assessment Framework for Privacy Policies

The framework consists of 16 parameters with all besides the first of them having up to four yes-no-questions. As already discussed, parameters are assessed towards a privacy score and a transparency score. The answer to each question is assessed and the awarded points sum up to a score in this parameter. Every parameter has a separate score. To balance the different number of each question, the score for each parameter is then normalized to be between 0 and 1. For questions that cannot be answered with yes or no (e.g. clicks needed) there was a table which assigned the clicks to points within this interval. Since convergence to the privacy-protective condition of the parameter raises the score, the score can be interpreted as "the higher the score, the better the privacy practices". Analogous, the transparency can be interpreted.

Agrawal et al. [19] weighted their categories with an importance factor, which is the case on the parameter level in this framework as well. Users can set a weighting factor for each parameter to operationalize their personal preferences. If the user is not able to come up with weights easily, the framework can also be used as a basis for an Analytic Hierarchy Process (AHP) like approach [26]. Hereby, the importance of every parameter is compared pairwise to each other and the result is a parameter importance ranking. However, with an increasing number of parameters, respondents might perceive this approach as exhaustive. For the remainder of this work the weighting factor was set to 1.

To make it easy for the user to see where a policy is positioned within the range of 100%, letters are assigned to relative scores. Therefore, we divided the range of possible scores into five quintiles such that a relative Privacy Policy Score (PPS) and respectively a relative Transparency Score (TS) with more than 80% get the best "A"-Ranking and the rankings with 20% and less get an "E"-Ranking which is the worst.

5.1 Parameters

The 16 parameters of the framework (cf. Table 1) cover different categories like accessibility, readability, the right to object, access, erasure and data portability. Whether the policy considers special treatment of children data and utilization of special data categories (Health, Race, Sex, ...) is covered as well. Also for the involvement of a third party, notification for changes or data breaches and notes on utilization for advertisement there are separate parameters. Due to space limitations, we are not able to describe each parameter and reasoning in detail, but for transparency each related GDPR article is noted in column § of Table 1.

5.2 Transparency Score

As shown in Table 1, all parameters are considered for the transparency score. Since it is modeled if the policy makes a statement, the value of a parameter question is 1 if the policy answered the question (irrespective how it was answered) and 0 if the question is not or contradictory answered.

Relative Transparency Score. The transparency score is based on the sum of the 16 parameters that each have a value between 0 and 1. The score for service i is calculated by formula 1 where $T_{i,j} \in \{0, 1\}$ represents the corresponding value of the parameters, and w_j is the weighting factor for parameter j. With $T_j^* = 1$ as the best possible score of parameter j, we get:

$$\text{Relative TS}_i = \frac{\sum_{j=1}^{n} w_j T_{i,j}}{\sum_{j=1}^{n} w_j T_j^*} = \frac{\sum_{j=1}^{n} w_j T_{i,j}}{\sum_{j=1}^{n} w_j} \tag{1}$$

5.3 Privacy Score

The privacy score needs a more distinct view on the parameters. Some parameters like the Flesch Reading Ease Score or if the policy is a multi-device policy can be assessed for all policies (cf. Table 1, sign: ✓). We did not consider the parameters marked with ✗ in Table 1, because some of them are not referring to the content of the policy, e.g. how easy it is to find the policy. Others do not necessarily need to be provided, e.g. the GDPR already states when a notification of policy changes needs to be provided. Gluck et al. [27] found contradicting signs: Despite that shorter notices are typically expected to be more effective, removing expected privacy practices from privacy policies sometimes led to less awareness of those practices, without improving awareness of the remaining practices. Thus, we decided not to consider these parameters for the privacy score.

However, there are also parameters which need to be stated (cf. Table 1, sign: 🔍⁻), e.g. the right of data portability, where we considered their absence negative for the privacy friendliness. In contrast, parameters which are in general not expected, but required if the service provider follows a certain practice (cf. Table 1, sign: 🔍⁺), e.g. transfer of data to third parties. Therefore, if no statement was given, we considered them to be positive for the privacy friendliness.

Table 1. The framework's parameters with their questions and how the parameters are considered for transparency (T) and the privacy friendliness of the policy (P).

#	Parameter Name	Parameter Description	T	P	§
1	Easily Acc. Form	1) Readability (Flesch Reading Ease Score)	✓	✓	12
2	Right to Object	1) Does the policy state a right to object? 2) Is an objection as easy as a consent?	✓	🔍	6, 7, 13, 21
3	Children	1) Is a binding age limit to use the service stated? 2) Is there a special policy for children? 3) Is there a mechanism to ensure that parents agree with the processing? 4) Does the policy state the procedure if children data has been processed unintentionally?	✓	🧍	8
4	Processing of Special Categories of Personal Data	1) Are special personal data categories processed? 2) Is it required contentwise for using the service? 3) Is there an explicit consent?	✓	🔍	9, 13
5	Necessary Information	1) Are identity and contact details of the controller stated? 2) Is a data protection officer stated? 3) Are the purposes of the processing for which the personal data are intended stated?	✓	🔍	13
6	Period of Storage	1) Is the storage period stated? 2) Are criteria determining the period stated?	✓	🔍	13
7	Right of Access	1) Is the right of access stated? 2) Is a fee charged?	✓	🔍	12, 13, 15
8	Right to Erasure	1) Is the right to erasure stated? 2) Is the time to fulfil the erasure request stated? 3) Period until fulfilment	✓	🔍	12, 13, 17
9	Data Portability	1) Is the right to data portability mentioned?	✓	🔍	13, 20
10	Third Countries	1) Is data processed in third countries? 2) Does the policy state these countries? 3) Is data transferred to countries with adequate level of protection (e.g. EU-U.S. Privacy shield)?	✓	🔍	45, 46, 47, 49
11	Data Breach Notification	1) Is a personal notification after a data breach explicitly stated? 2) Period until notification	✓	✗	34
12	Third Parties	1) Is a third party involved by design? 2) Does the policy state who the third party is? 3) Does the policy explicitly state the purpose? 4) Is the scope of the transferred data stated?	✓	🔍	13
13	Search for the Policy	1) Is there a link on the homepage that leads to the policy for the device quickly? 2) How many clicks are needed from the homepage to find the link to the policy?	✓	✗	12, 13
14	Change Notificat.	1) Is there a notification after policy changes?	✓	✗	13
15	Special Device Policy	1) Is the present policy a multi-policy? 2) Is it clear, the policy is for the IoT product?	✓	✓	
16	Lifecycle	1) Can information stored on the device be deleted?	✓	🔍	

✓: Used, ✗: Not used, 🔍/🔍: If not present, rated positive/negative, 🧍: Only for toys

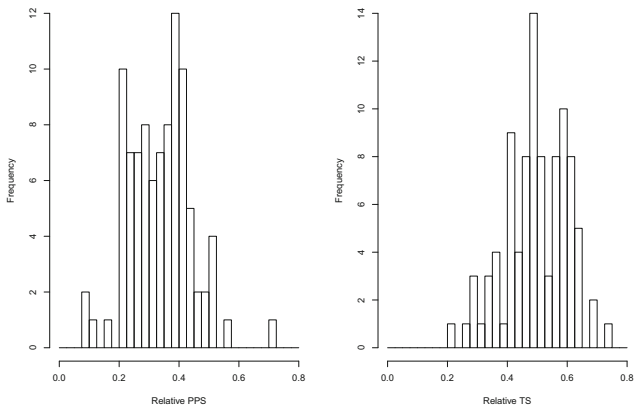


Fig. 1. Histogram of PPS and TS of examined policies

The parameter marked with 🧍 should only apply to devices which are used by children. Since for many devices there is no clear statement of the target audience, we considered it only for toys.

Relative Privacy Policy Score. The value which enables comparisons along different policies is called relative Privacy Policy Score (relative PPS). The relative PPS for service i is calculated by formula 2 where j is the parameter id, x_j is the weighting factor for parameter j, $P_{j,i}$ is the score of parameter j for Service i and with $P_j^* = 1$ as the best possible score of parameter j, we get:

$$\text{Relative PPS}_i = \frac{\sum_{j=1}^{n} x_j P_{i,j}}{\sum_{j=1}^{n} x_j P_j^*} = \frac{\sum_{j=1}^{n} x_j P_{i,j}}{\sum_{j=1}^{n} x_j} \tag{2}$$

6 Results

A set of 113 IoT devices was created, but while collecting policies we found three products without a policy which would be ranked with 0% in both dimensions. For legibility reasons we removed these ones and ended up with 110 products to assess. They were divided into three umbrella categories Smart Home, Smart Health and Toys, which are subdivided in groups e.g. Thermostat, Light, Washer, etc. Some privacy policies covered multiple devices or they were a privacy policy for all of the company's services. According to the assessment framework in Sect. 4.3, privacy policies were assessed and ranked based on their achieved privacy and transparency scores. In the end, we assessed 94 policies: 14 policies covered 30 devices and 80 policies were for a single IoT device. Two devices changed their policy during the assessment period.

6.1 Ranking Results

Table 2 shows the results of the privacy and transparency scores grouped into the respective subgroups. Figure 1 presents histograms for the relative privacy policy respectively transparency score.

Table 2. Summary statistics of examined policies

Area	Subarea	#	PPS Score					Rel. PPS (%)		Transparency					Rel. TS (%)	
			A	B	C	D	E	Mean	STD	A	B	C	D	E	Mean	STD
Smart Home	Coffee Machine	5	0	0	1	4	0	31.67	8.39	0	0	4	1	0	47.50	10.37
	Light	5	0	0	2	3	0	35.56	8.67	0	1	4	0	0	53.75	6.04
	Security	9	0	0	3	5	1	32.80	11.36	0	1	7	1	0	48.61	9.80
	Thermostat	6	0	0	3	3	0	36.69	11.10	0	1	4	1	0	50.43	11.35
	Washer	5	0	1	2	2	0	37.91	20.83	0	1	3	1	0	54.17	12.68
	Others	28	0	0	7	21	0	34.71	8.95	0	5	20	3	0	50.52	8.99
	Total	58	0	1	17	38	2	34.70	10.50	0	9	42	7	0	50.55	9.37
Health	Fitness Tracker	7	0	0	2	5	0	36.11	6.39	0	1	6	0	0	53.72	4.91
	Scale	15	0	0	1	12	2	28.75	11.56	0	3	6	6	0	43.89	12.93
	Others	5	0	0	1	4	0	33.89	8.22	0	1	4	0	0	52.29	6.93
	Total	27	0	0	4	21	2	31.61	10.14	0	5	16	6	1	47.99	11.18
	Toy	9	0	0	3	6	0	34.05	12.66	0	2	6	1	0	50.92	13.18
∑	Total	94	0	1	24	65	4	33.75	10.59	0	16	64	14	0	49.85	10.26

6.2 Statistics on the Privacy Policies

The results do not appear to have similarities with a normal distribution. We conducted a Shapiro-Wilk-Test [28] to confirm or reject this hypothesis. It is a high-quality test for normal distribution that can be applied on relatively small samples. The p-value predicts how likely it is to get such results from a normal distribution. With a p-value of 0.1368 for the relative PPS and p-value of 0.3146 for the relative TS, we assume that the distribution of the privacy scores and the distribution for the transparency score are not close to a normal distribution.

Due to the results of Gluck et al. [27], we were also interested in the relationship between the length and the privacy respectively transparency score of the privacy policies. Since the plots (cf. Fig. 2) show some clusters, we conducted Spearman correlation tests [29]. For the correlation between the number of words in the policy and the privacy score we found a moderate effect size ($\rho_{PPS} \approx 0.518$ with p-value $\approx 8.8 \cdot 10^{-8}$). Analogous, for the correlation between the number of words in the policy and the transparency score we found a strong effect size ($\rho_{TS} \approx 0.723$ with p-value $\approx 2.2 \cdot 10^{-16}$. Both correlations are statistically highly significant and allow us to conclude that there is a relationship between the length of the policy and the privacy respectively transparency score.

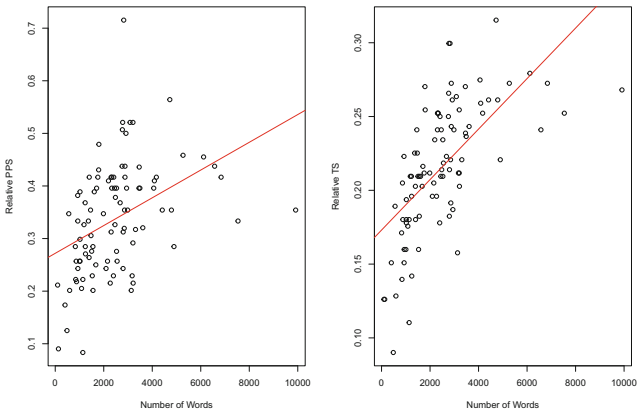


Fig. 2. Relationship between length and relative PPS/TS

7 Discussion

The ranking of the both scores within the quintiles shows that none could get an A-rating. This might improve when the GDPR is put in place in May 2018. However, being compliant to the GDPR could also mean to inform about certain privacy practices without them being more privacy friendly. Difficulties in finding the right policy raises also the question whether companies use privacy policies to inform the users or if they just use them as a legal cover.

The result of the correlation between scores and length should not be misunderstood as a motivation to provide longer policies because longer policies seem to be better. More likely, the result is due to the fact that in longer policies more topics can be covered. We expect a certain length where this effect will invert.

7.1 Limitations and Threats to Validity

Despite all care, the assessment framework cannot replace the detailed analysis of a lawyer. Although, the questions are Additionally, it was not possible to test the implementation of the policy. All assessment is based on the written policy and it is not guaranteed that companies follow their own rules. Future research should crosscheck contents and execution of the policy. Labels like TRUSTe, which the FTC approach took into account for a measure of enforcement [18], can be an indicator that their policies indeed reflect their practices. Nevertheless, even for labels like TRUSTe, there is reason for critique e.g. in meaningfulness [30].

We only examined English privacy policies. We can not exclude that the policies' contents differ between the different language versions. According to Article 12 of the GDPR the policy must be provided "in a concise, transparent, intelligible and easily accessible form, using clear and plain language". The

availability of a language other than English is not explicitly mentioned in the GDPR but the line of argument could be that this supports the requirements.

A weak point of parameter 13 (Search for the Policy) is that the effort to find a policy is not a reliable measure because it is dependent on who looks for it. Some companies use the same policy for their products as for their websites and some companies don't declare the range of application which makes it difficult to ensure that the present policy is the right one for the IoT product. However, we could statistically show that there was no learning effect when searching for the policy since the number of steps was not significantly lower at the last investigated policies.

7.2 Future Extension of the Framework

One design goal of this framework was its openness to extensions. New parameters can be easily added, the utilization of a relative score instead of an absolute score makes allowance for this, because it allows a step-wise re-assessment. One can easily think of further requirements for a good privacy policy/practice which is not considered in this framework yet, but future work could create new parameters to operationalize them. We list some of the additional parameters, we also considered, assessed but not included in the final version of the framework. Procedure of data sharing after a *corporate merge or bankruptcy*. Has the parent company access to personal information after a merge? We didn't include this parameter in the final framework, because we couldn't find a statement how reliable this declaration would be if there would really be a merge or bankruptcy. A parameter considering the data processing if *the user is not the owner*, but e.g. a guest in a smart home where microphones listen for commands and listen to the guests, who have not given consent [31]. Is the scenario of an incidental use considered? Are there mechanisms to protect against an incidental use? Since as of today, this seems to be a non resolved issue, we also did not consider this parameter in our framework. For the same reason, we did not consider *interacting systems*, where each system has its own privacy policy and there is a chance of inconsistencies arising when systems work together.

8 Conclusion and Future Work

This paper presents an extendable assessment framework for privacy policies consisting of 16 parameters. We collected 94 privacy policies covering 110 devices. Users can look up certain topics or compare devices according to their own preferences.

The results of this comparative study show that most of the examined privacy policies of IoT devices/services are insufficient to address the GDPR requirements and beyond. Many topics are currently not addressed in privacy policies but will need to be covered until May 2018, when the GDPR comes into effect.

Difficulties in finding the right policy raises the question whether the purpose of privacy policies is to inform the users and make them conscious of the data

processing or if it is just a legal cover, which deserves further research. The transparency dimension tried to operationalize this aspect but further development and improvement of this dimension is required.

During the analysis of this work it also seemed as though that products on the European market have fewer functionalities than US products. Some devices are not even available for EU citizens, perhaps due to the higher requirements of European law. Future work could check this impression. Additionally, there might be differences in the cont the same policies in different languages and future research should include a comparison.

To make people more aware about the shortcomings of privacy policies, a public ranking website should be designed. Based on the current framework users could set the privacy preferences and a personalized score could be calculated. Awareness for privacy topics might help to force companies to reform their practices. To avoid manually processing a larger number of policies, an automatic assessment tool could be designed and developed, e.g. based on a machine learning approach. In particular, we aim at extending the framework by using the assessed privacy policies as corpus and building predictive models using machine learning and natural language techniques. Furthermore, considering semantic features of privacy policies could result in analyzing and benchmarking IoT privacy policies with high accuracy. Such automatic and adaptive models coupled with usable and informative user interfaces can be helpful to support users in analyzing and retracing the data processing practices of IoT services they intend to subscribe.

Acknowledgments. This research was partly funded by the German Federal Ministry of Education and Research (BMBF) with grant number: 16KIS0371.

References

1. Stankovic, J.A.: Research directions for the internet of things. IEEE Internet Things J. **1**(1), 3–9 (2014)
2. Information Commissioner's Office: Privacy regulators study finds Internet of Things shortfalls (2016)
3. Mayer, C.P.: Security and privacy challenges in the internet of things. In: Electronic Communications of the EASST, vol. 17 (2009)
4. DZone: The DZone guide to Internet of Things (2016)
5. Milne, G.R., Culnan, M.J.: Strategies for reducing online privacy risks: why consumers read (or don't read) online privacy notices. J. Interact. Mark. **18**(3), 15–29 (2004)
6. European Commission: Special Eurobarometer 431: Data Protection Report (2015)
7. Jensen, C., Potts, C., Jensen, C.: Privacy practices of internet users: self-reports versus observed behavior. Int. J. Hum.-Comput. Stud. **63**(1–2), 203–227 (2005)
8. Casadesus-Masanell, R., Hervas-Drane, A.: Competing with privacy. Manag. Sci. **61**(1), 229–246 (2015)
9. Xia, F., Yang, L.T., Wang, L., Vinel, A.: Internet of things. Int. J. Commun. Syst. **25**(9), 1101–1102 (2012)

10. European Parliament, Council of The European Union: Regulation (EU) 2016/679 General Data Protection Regulation (GDPR) (2016). http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32016R0679. Accessed 15 Jan 2018
11. European Commission: Proposal for a Regulation on Privacy and Electronic Communications (ePrivacy Regulation) (2017). http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52017PC0010. Accessed 15 Jan 2018
12. European Interactive Digital Advertising Alliance (EDAA): The e-privacy regulation - good or bad for european consumers? (2018) http://www.likeabadmovie.eu/. Accessed 15 Jan 2018
13. Engeler, M., Felber, W.: Draft of the ePrivacy Regulation from the perspective of the regulatory practice (2017). http://rsw.beck.de/rsw/upload/ZD/ZD_Sonderveroffentlichung_Engeleer_Felber_engl.pdf. Accessed 15 Jan 2018
14. Pellikan, L.: Bundesregierung: ePrivacy-Verordnung kommt erst 2019. W&V of 22 November 2017 (2017). https://www.wuv.de/digital/bundesregierung_eprivacy_verordnung_kommt_erst_2019. Accessed 15 Jan 2018
15. Ziegeldorf, J.H., Morchon, O.G., Wehrle, K.: Privacy in the Internet of Things: threats and challenges. Secur. Commun. Netw. **7**(12), 2728–2742 (2014)
16. Smith, H.J., Milberg, S.J., Burke, S.J.: Information privacy: measuring individuals' concerns about organizational practices. MIS Q. **20**(2), 167 (1996)
17. Milne, G.R., Culnan, M.J.: Using the content of online privacy notices to inform public policy: a longitudinal analysis of the 1998–2001 U.S. web surveys. Inf. Soc. **18**(5), 345–359 (2002)
18. Peslak, A.R.: Internet privacy policies. Inf. Resour. Manag. J. **18**(1), 29–41 (2005)
19. Agrawal, R., Grosky, W.I., Fotouhi, F.: Ranking privacy policy. In: IEEE 23rd International Conference on Data Engineering Workshop, pp. 192–197 (2007)
20. Flesch, R.: A new readability yardstick. J. Appl. Psychol. **32**(3), 221–233 (1948)
21. Ranking Digital Rights: 2017 Corporate Accountability Index (2017)
22. Terms of Service; Didn't Read project: Website (2017). https://tosdr.org/. Accessed 15 Jan 2018
23. Zimmeck, S., Bellovin, S.M.: Privee: an architecture for automatically analyzing web privacy policies. In: Proceedings of the 23rd USENIX Security Symposium, 20–22 August 2014. USENIX Association (2003)
24. Zimmeck, S., et al.: Automated analysis of privacy requirements for mobile apps. In: NDSS 2017 Network and Distributed System Security Symposium (2017)
25. WebpageFX: Readability Test Tool. https://www.webpagefx.com/tools/readable/. Accessed 15 Jan 2018
26. Saaty, T.L.: What is the analytic hierarchy process? In: Mitra, G., Greenberg, H.J., Lootsma, F.A., Rijkaert, M.J., Zimmermann, H.J. (eds.) Mathematical Models for Decision Support, pp. 109–121. Springer, Heidelberg (1988). https://doi.org/10.1007/978-3-642-83555-1_5
27. Gluck, J., et al.: How short is too short? Implications of length and framing on the effectiveness of privacy notices. In: Symposium on Usable Privacy and Security (SOUPS) (2016)
28. D'Agostino, R.B., Stephens, M.A., (eds.): Goodness-of-Fit Techniques, Volume 68 of Statistics, 5. print edn. Dekker, New York (1986)
29. Hollander, M., Wolfe, D.A.: Nonparametric Statistical Methods, 2nd edn. Wiley-Interscience (1999)
30. McCarthy, J.: TRUSTe decides its own fate today - slashdot (1999)
31. von Leitner, F.: Das IoT-Problem (2017). https://ptrace.fefe.de/iot. Accessed 15 Jan 2018

C.6 Applying Privacy Patterns to the Internet of Things' (IoT) Architecture

Sebastian Pape and Kai Rannenberg. Applying privacy patterns to the internet of things' (IoT) architecture. *Mobile Networks and Applications (MONET) – The Journal of SPECIAL ISSUES on Mobility of Systems, Users, Data and Computing*, 24(3):925–933, 06 2019. doi: 10.1007/s11036-018-1148-2. URL https://doi.org/10.1007/s11036-018-1148-2

Mobile Networks and Applications (2019) 24:925–933
https://doi.org/10.1007/s11036-018-1148-2

Applying Privacy Patterns to the Internet of Things' (IoT) Architecture

Sebastian Pape[1] · Kai Rannenberg[1]

Published online: 2 October 2018

Abstract
The concept of cloud computing relies on central large datacentres with huge amounts of computational power. The rapidly growing Internet of Things with its vast amount of data showed that this architecture produces costly, inefficient and in some cases infeasible communication. Thus, fog computing, a new architecture with distributed computational power closer to the IoT devices was developed. So far, this decentralised fog-oriented architecture has only been used for performance and resource management improvements. We show how it could also be used for improving the users' privacy. For that purpose, we map privacy patterns to the IoT / fog computing / cloud computing architecture. Privacy patterns are software design patterns with the focus to translate "privacy-by-design" into practical advice. As a proof of concept, for each of the used privacy patterns we give an example from a smart vehicle scenario to illustrate how the patterns could improve the users' privacy.

Keywords Privacy by design · Cloud computing · Fog computing · Internet of things · Privacy patterns · Autonomous cars · Smart vehicles

1 Introduction

With an estimated number of 50 billion ubiquitous and interconnected devices by the year 2020 the Internet of Things (IoT) is growing rapidly [1]. Since its beginning, the IoT concept has been relying on a strong computing infrastructure built on cloud computing services [2]. However, new concepts and technologies to manage the huge amount of devices are gaining importance. The backbone evolved into a more heterogeneous concept which is known as fog (or sometimes mist or edge) computing. A literature survey by Thien and Colomo-Palacios [3] showed that the main purposes or developments of the architecture addressed six different areas: resource management, energy efficiency, offloading, data processing, performance enhancement and networking. All of these are merely performance problems.

However, privacy concerns in the IoT are not only a research topic [4], but have arrived at customers which were spied by their devices [5, 6]. Adams [7] notes that due to the nature of IoT devices and the way they collect information, their use leads to a higher risk of having information collected and shared. Often the IoT devices and sensors come together with mobile apps. Papageorgiou et al. [8] discovered in the mobile health domain that most of the apps do not follow well-known practices and guidelines jeopardizing the privacy of millions of users. Weinberg et al. add that in the IoT environment the user faces a trade-off between convenience and privacy [9]. Moreover, Adams [7] and Walker [10] found that the regulators cannot keep up with the advances in the market, e.g. because of the speed with which data is exchanged. Apparently, privacy notices or policies could reduce the risk of disclosing personal information, but customers got increasingly frustrated with them [11, 12]. Since this discovery, not much has changed, as a recent study on IoT privacy policies shows [13].

We argue that in particular with the *General Data Protection Regulation* (*GDPR*) which has just become effective, more emphasis should be put on designing privacy-friendly services (privacy by design). Therefore, we investigate how the different characteristics within the IoT / Cloud / Fog architecture could be used to improve users' privacy.

The remainder of this work is organized as follows. Section 2 gives a brief introduction into fog computing and describes related work, in particular about privacy in IoT environments and privacy patterns. In Section 3 suitable privacy patterns are mapped to the IoT / Cloud / Fog architecture.

✉ Sebastian Pape
sebastian.pape@m-chair.de

Kai Rannenberg
kai.rannenberg@m-chair.de

[1] Deutsche Telekom Chair of Mobile Business & Multilateral Security, Goethe University Frankfurt, Theodor-W.-Adorno-Platz 4, 60323 Frankfurt, Germany

Section 4 gives some examples using scenarios of smart vehicles for (partially) autonomous driving. Section 5 discusses the findings and concludes this work.

2 Background and related work

In this section, we first briefly sketch the differences between cloud and fog computing and how they work together with IoT devices. Next, we describe work on privacy for IoT systems including relevant work on fog and cloud computing when appropriate. Since our work strongly relies on it, we also address research on privacy patterns.

2.1 Fog computing conceptual model

Our description of the conceptual model for fog computing follows the respective NIST special publication [14]. The idea of *cloud computing* was to have central large datacentres with huge amounts of computational power. However, it has been shown that with the exponential growth of IoT devices and the amount of data they produce this architecture produces costly, inefficient and in some cases infeasible communication [15]. This is in particular true for services with low latency requirements, e.g. real-time interactions. In order to achieve minimal latency and reduce costs, a new architecture with distributed computational power closer to the IoT devices was developed – *fog computing* (cf. Fig. 1). In this architecture, a substantial amount of data processing is done in decentralised, distributed nodes and thus complementing the centralized cloud computing model when serving IoT devices. According to the NIST report [14], no clear distinction between the names *fog computing*, *edge computing*, *mist computing* or *cloudlets* exists. However, following Bonomi et al. [16] *edge computing* is the underlying principle which allows data storage and computation at the edge of the network, and thus close to the end users.

Figure 1 shows the three-layer service delivery model, where fog nodes reside between the IoT devices and the cloud service. Fog computing is not a replacement but an extension to the cloud computing architecture [17]. Naturally, the fog nodes are context-aware, e.g. they know about their location. Fog nodes can be clustered vertically to allow isolation or horizontally to support federation. According to NIST [14] and Thien and Colomo-Palacios [3], fog computing has the following essential characteristics:

- *Contextual location awareness, and low latency*: Since the fog nodes are closer or often even co-located to the IoT devices, responses by these nodes can be delivered faster than by the centralised cloud computing system. Natively, the nodes know about their logical location in the context of the overall computational system.

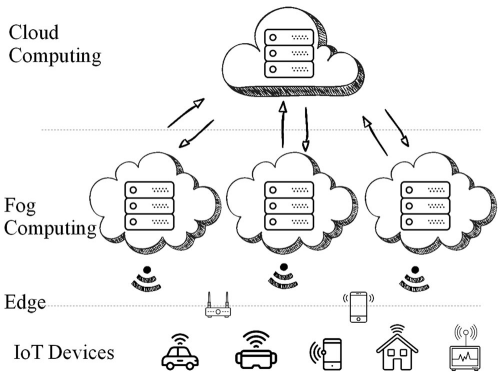


Fig. 1 Three-layer service delivery model

- *Geographical distribution*: In contrast to the centralised cloud computing paradigm, fog nodes are widely distributed and geographically identifiable. This is necessary to provide services for example to vehicles. The fog computing nodes can be distributed along the track the vehicle is moving on.
- *Heterogeneity*: In contrast to cloud computing, where there is only one large node, fog computing nodes can consist of different forms and types of computing nodes.
- *Interoperability and federation*: In order to achieve seamless service distribution, the cooperation of different providers is needed.
- *Real-time interactions*: Natively, applications which involve real-time interactions make use of fog computing while traditional batch processing can still be performed in cloud computing services.
- *Scalability and agility of federated, fog-node clusters*: New clusters of fog computing nodes can easily be added or existing clusters can be extended.
- *Edge analytics*: Fog computing can support analysing data locally instead of sending it to the cloud for analysis.

2.2 Privacy in the IoT

Kumar and Patel [18] give a very high level overview of privacy concerns in the IoT. They build the four groups of privacy in the device, during communication, in storage and at processing. Analogous, Martinez-Balleste et al. [19] identify privacy threats in Smart Cities and group them into the five dimensions: identity privacy, query privacy, location privacy, footprint privacy, and owner privacy. In a next step they point to technologies that could address these threats.

Kowatsch and Maass [4] addressed privacy concerns and acceptance of IoT services from the perspective of information systems. They proposed and tested an instrument to evaluate

IoT services by extending the privacy calculus model [20] and combining it with the Technology Acceptance Model [21]. Their goal was to gain insights about the users' willingness to share information to use IoT services to provide recommendations to policy makers and developers how to design privacy-aware IoT services.

Kozlov et al. [22] discuss security and privacy threats in the IoT architecture and also connect them to EU legislation. They do not mention the cloud or edge computing paradigms, but have a very similar architecture where they elaborate on privacy and security threats. One of their conclusions is that many threats are similar to those in already existing architectures. Complementarily, Lee et al. [23] focus on security issues in the fog computing supported IoT cloud and argue that its adoption introduces several unique security and privacy threats. Stojmenovic et al. [24, 25] studied issues such as security, demand response, privacy, fault tolerance in the context of fog computing. They in particular focus on man-in-the-middle attacks and sketch how to adapt a data aggregation scheme from Lu et al. [26] to address privacy issues. In their extensive work, which is focused on security threats, Ni et al. [27] also list some privacy threats along with discussing security and privacy requirements and state-of-the-art solutions in fog computing. Tayeb et al. [28] and Sadeghi et al. [29] focus on an industrial viewpoint and discuss security threats and challenges separately for all the layers. They point out that industrial systems are an attractive target since they generate, process and exchange vast amounts of security-critical and privacy-sensitive data. This way they show that security and privacy are often two sides of the same coin. Yi et al. [30] highlight privacy issues in data privacy, usage privacy, and location privacy on the new aspects of fog computing by surveying the literature.

However, only few of these works propose approaches on how to address privacy issues. Those who do, rarely make use of the specific architecture of fog computing to improve the users' privacy. Closest to our work is the work from Rahman et al. [31] who discuss and compare IoT programming frameworks in order to give some guidance to find the most suitable. For that purpose, Rahman et al. define a taxonomy to classify the architecture which makes essential architectural aspects explicit in order to compare the aspects' influence on functional properties. Among other features, privacy issues are also discussed. Naturally, the decision for a programming framework is on a different level than the application of privacy patterns in the IoT architecture. Thus, the guidance points in the same direction but towards different levels of abstraction compared with our work.

2.3 Privacy patterns

Patterns are a useful method – often used in software design – to describe already known solutions and best practices for design problems [32]. Yoder and Baraclow were the first who developed patterns to address information security issues [33].Based on the Common Criteria [34] Schumacher identified two user-focused privacy patterns [35]. Privacy patterns can be considered to be a subset of design patterns with the focus to translate "privacy-by-design" into practical advice for software engineering [36].

There have been contributions to privacy patterns since the beginning of this century, although some of them do not include the term privacy pattern. Schümmer introduces six patterns and groups them into the two categories: blocking personal information from being transmitted from the user and filtering information sent from others to the user [37]. Romanosky et al. [38] identify three privacy patterns for web-based activity. Graf et al. [32] describe the development of User Interfaces Patterns for Privacy Enhancing Technologies (PET).

Doty and Gupta [39] note a lack of concrete guidance for implementing Privacy-by-Design. Therefore, they proposed privacy design patterns adapted from software engineering design patterns and established a site to allow a collaborative collection and development of privacy patterns [36]. The privacy patterns website aims to standardize the language for privacy-preserving technologies, to document common solutions to privacy problems and to help designers identify and address privacy concerns. In a similar manner, the Privacy Design Pattern Library Website [40] provides a pattern library for making privacy policies understandable. Both libraries [36, 40] provide a database where common patterns can be looked up and searched.

3 Mapping privacy patterns to architecture considerations

Ni et al. [27] list four aspects of information which is privacy relevant in the IoT.

- *Identity Information*: Any information that can link to a specific user, e.g. name, address, telephone number, credit card number or public-key certificate.
- *Data*: Various sensitive information, such as a user's preferences, occupation, health status and political inclination.
- *Usage Information*: Usage pattern with which a user utilizes the services offered, e.g. the readings of a smart meter.
- *Location Information*: With location information an attacker is able to identify a user's trajectory, identity, points of interest. It seems that location privacy is a kind of privacy that we have to sacrifice to use online services, such as navigation and location-based services.

In the following subsections, we first introduce a privacy pattern (if possible) based on the website privacypatterns.org [36] and then show how it can be applied to the IoT with the cloud / fog computing architecture behind. We only discuss privacy patterns where the characteristics of the specific IoT / cloud computing / fog computing architecture can be exploited.

3.1 Personal data store

The main idea of the "Personal Data Store" privacy pattern is that users keep control over their personal data and store it on a personal device. The pattern can only be applied for data produced by the user and not for data produced by a third party. The pattern aims to prevent the user to lose control over their data when submitting it to a server operated by a third party or storing it there.

For IoT devices this could mean that identity information or data is stored locally and (if possible) computations are also done locally (see Fig. 2). If the IoT device is too small and has too little computational power, a workaround would be to make use of the user's mobile phone. Many devices connect to the Internet via the user's phone or they use the user's phone with an app as interface to control the device. Often the data is stored in the cloud and accessed by the phone's app. If this cannot be changed, a possibility would be that the IoT device encrypts the data and the decryption key is on the mobile phone while the cloud respectively fog nodes do not have the decryption key. Nowadays, many mobile phones offer a decent level of computational power.

3.2 Data isolation at different entities

The main idea of the "Data Isolation at Different Entities" privacy pattern is that if data or usage information is distributed among several entities, all of the entities can only see a part of the data. This improves the users' privacy since it gets harder to profile him/her.

In the determined architecture, the fog nodes or clusters would be an excellent layer to enforce isolation. As already stated, if fog nodes are clustered vertically, each cluster can belong to a different organisation respectively provider (see Fig. 3). Note that this privacy pattern does not prevent collusion attacks. Several providers could exchange information to profile a single user or a group of users. Since the unauthorized exchange of data does not need to use the IoT infrastructure, and thus cannot be controlled, the easiest way to prevent it are legal arrangements. This pattern can be easily combined with the following patterns "Decoupling Content and Location Visibility", "Added Noise Measurement Obfuscation", and "Data Aggregation".

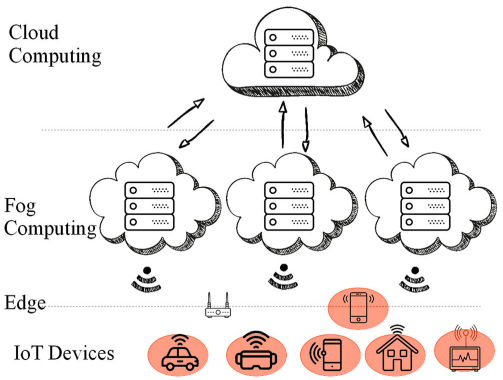


Fig. 2 Personal Data Store at IoT Devices or the Mobile Phone

3.3 Decoupling content and location information visibility

Users often share content in socially oriented services. Since many consumer devices, e.g. mobile phones, have location data available, applications may attach location information when uploading data. However, in fog computing environments it is difficult to protect the users' locations as users normally access the services provided by the nearest fog node. This node can then assume the user is nearby, and thus infer about the users' location [27]. On the other hand, it would be possible that each fog node can specifically monitor if a user is transmitting location information and then either make the user aware or remove this information (see Fig. 4). Additionally, if users access the same service at multiple fog nodes their movement can be disclosed. This can be countered by vertically clustering the fog nodes as already discussed in the previous subsection.

3.4 Added noise measurement obfuscation

If users repeatedly use a resource over time, detailed measurement may reveal further information about the users such as personal habits. This privacy pattern suggests to add some noise to the measurements which cancels itself in the long term.

If the fog nodes are run by a provider the user trusts they can add the noise, so that the cloud service provider only gets the noisy data (see Fig. 4). Users who do not trust the fog computing provider, could do the same on their mobile phone if the IoT device connects through it to the fog or the cloud computing service. However, the implementation depends a lot on the type of usage information and the purpose of its collection. A provider probably wouldn't allow the users to add noise by themselves if the information is used for billing

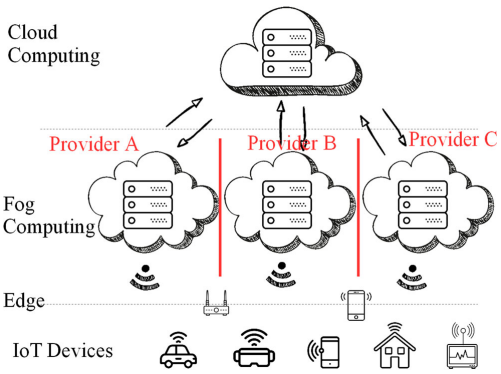


Fig. 3 Data isolation at different entities

purposes. Additionally, a there is trade-off between the usefulness of the data and the protection of the user's privacy: the more noise is added, the less useful the data is, but the better the user's privacy is protected.

3.5 Aggregation of data

Analogous to the previous privacy pattern, where noise was added, another possibility to prevent the leakage of further information is the aggregation of data. For example the usage information of multiple users or the usage information of a single user aggregated over time.

Analogous to the adding of noise, the aggregation can either be done by a trustworthy provider or (in the case of aggregation over time) by the users themselves (see Fig. 4). The same restrictions and considerations apply. However, depending on the purpose, the aggregation can be done with homomorphic encryption as described in the next subsection.

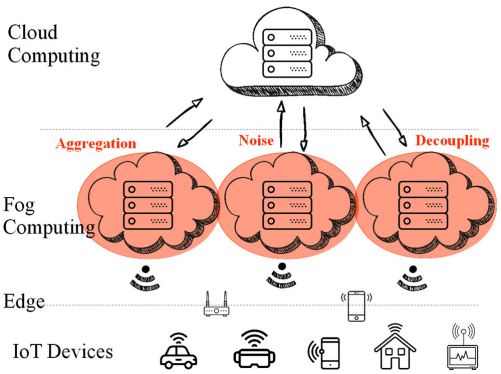


Fig. 4 Decoupling, noise adding and aggregation of data at the fog nodes

3.6 Aggregation gateway

The "Aggregation Gateway" privacy pattern is useful, if and when a service provider needs a continuous measurement and adding noise is not acceptable. This problem can be solved by using homomorphic encryption (e.g. Paillier [41]) and a trusted third party aggregating the measurements of multiple users.

Each measurement is encrypted by the IoT device or the user's mobile phone. The key is shared between the fog node and the IoT devices, e.g. by making use of Shamir's Secret Sharing Scheme [42]. The encrypted measurements from a group of users are transmitted to the cloud computing provider. Since the data passes through a fog computing node, which may have the decryption key, an additional encryption, e.g. transport layer security (TLS) [43] needs to be applied. Since the measurements have been encrypted with a homomorphic encryption system, the cloud is able to operate on the data and can, e.g. aggregate it, without being able to access the data in clear. The result of the computation, e.g. the aggregation, can then be sent to the fog node. With the previously shared secret, the fog node can decrypt the result and access it in clear without learning about the individual values of the different users or devices (see Fig. 5). It is worth mentioning that homomorphic encryption in general has additional computational costs, but the aggregation operation when applying this privacy pattern to the IoT architecture is located at the party with the most computational power. For a state-of-the-art scheme, we refer the reader to recent work from Okay and Ozdemir [44].

3.7 Single Point of Contact

With distributed storage, a specialised privacy management becomes necessary, the "Single Point of Contact". The Single Point of Contact should be able to issue security tokens, authenticate local domain users as an Identity Service Provider, certify attributes as an Attribute Provider, and accept external claims as a Relying Party.

The cloud computing service can manage and coordinate the storage on different fog nodes by providing the services described above (see Fig. 6).

4 Evaluation of the patterns by applying the privacy patterns in the smart vehicles scenario

In order to demonstrate the applicability of privacy patterns, we also show how they could be applied in a typical IoT / Fog Computing scenario. Thien and Colomo-Palacios [3] found five relevant scenarios in the literature survey about fog computing: healthcare, smart grid, smart vehicles, urgent computing (e.g. disaster support) and augmented reality. For all those

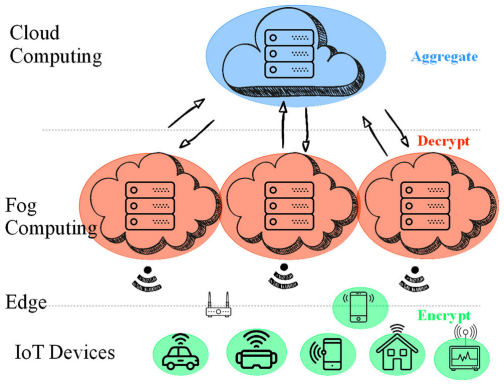


Fig. 5 Aggregation gateway

scenarios, one can easily come up with a connection to IoT (sensors). In order to have one coherent scenario for all seven patterns, we chose smart vehicles to demonstrate how the privacy pattern can be applied in practice. For that purpose, we build on former work from Rannenberg [45] which investigates privacy issues in smart vehicles, especially in relation to autonomous driving. We briefly sketch the scenario in the next subsection before we apply the privacy patterns.

4.1 The "smart vehicles" scenario

"Smart Vehicles" describes the automation of vehicles to support the driver and often involves the use of artificial intelligence. The support of the driver ranges from warnings through driving assistance, via automation of some tasks to fully autonomous driving. For a systematic evaluation, the standard J3016 from SAE provides six levels with a detailed description of each automation level [46]. Rannenberg [45] notes that an autonomous car relies much more on interaction with the outside world than a human-driven car. This raises privacy concerns and motivates Rannenberg to analyse data flows and corresponding privacy impact. For that purpose, Rannenberg defines four use cases. The first two of them are sufficient to apply the privacy patterns discussed in Section 3.

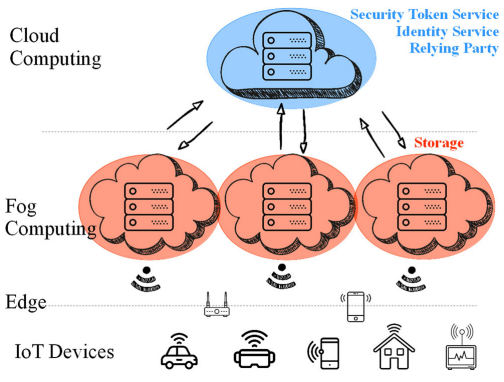


Fig. 6 Single point of contact

4.1.1 Use Case 1: Interstate pilot using driver for extended availability

The driving robot takes over the driving task, but only on interstates or interstate-like expressways. During autonomous journeys, drivers become passengers who can take their hands off the steering wheel and their feet off the pedals and pursue other activities. The driving robot coordinates a safe handover to the driver and may even stop the car at a safe place if needed. We assume for our application of the privacy patterns that there is a fog computing infrastructure along the interstate.

4.1.2 Use Case 2: Autonomous valet parking

The driving robot parks the vehicle at a nearby or remote location after the users have exited and cargo has been unloaded. Later the driving robot drives the vehicle from the parking location to a desired destination. The driving robot re-parks the vehicle. The driver saves the time of finding a parking spot as well as of walking to/from a remote parking spot. In addition, access to the vehicle is eased (spatially and temporally). Moreover, parking space is used more efficiently and search for parking is arranged more efficiently. We assume for our application of the privacy patterns that there is a fog node at each parking location.

4.2 Application of the privacy patterns

For each of the privacy patterns discussed in Section 3, we show how they would be applied to the scenario discussed above to achieve a more privacy friendly design of services.

4.2.1 Personal data store

The idea of the personal data store is that information is not stored in a central database but under the control of the user. Rannenberg already argues that for "data stored in a car are under the sole control of the car's owner or driver, [...] determining responsibility for these data may be relatively easy" [45]. This holds for several of the scenarios for smart vehicles and is in line with the privacy pattern of "Personal Data Storage".

In Use Case 2, traffic control centres or other entities involved in the choice of parking spaces should not ask the drivers or passengers for all kinds of priorities for a parking space or route, but instead give some options, so that the user

or a local system assisting the user, can choose. This reduces the risk of a centralized processing of users' attitudes with regard to prices and locational preferences [45, p., 513].

4.2.2 Data isolation at different entities

The idea of data isolation at different entities is to avoid building full profiles on the user and restrict each entity to only a part of the data. In Use Case 1, with different fog clusters along the interstate, the route of the vehicle, respectively user cannot be tracked that easily – if the fog clusters belong to different entities. In Use Case 2, the driver might have different preferences and habits at different locations. By isolating this data, building profiles is made more difficult.

4.2.3 Decoupling content and location information visibility

The idea of decoupling content and location information visibility is to avoid that one entity learns characteristics about the user along with his or her location. In Use Case 1, the manufacturer of the car might be interested in some usage statistics of the car. However, there is no need that the manufacturer learns the location information. In Use Case 2, the location cannot be hidden, thus the aim would be a minimisation of all other data collected at the fog responsible for coordinating the parking.

4.2.4 Added-noise measurement obfuscation

The idea added-noise measurement obfuscation is to hide certain characteristics by blurring the data. In Use Case 1, traffic and congestion analysis does not need to identify individual cars or even drivers. For that purpose it can be helpful to add noise to the data in order to hide certain characteristics of the car, e.g. maximum acceleration, which might lead to an identification of the car and thus reduce set of possible cars and respectively drivers and owners. In Use Case 2, the exact location of the car might be blurred when sending requests for free parking spaces.

4.2.5 Aggregation of data

The idea of data aggregation is to not allow a certain entity to see single data. Analogous to the previous application in Use Case 1, in order to hinder the identifiability of individual cars, for traffic and congestion analysis it may be sufficient to work with aggregated values.

4.2.6 Aggregation gateway

An aggregation gateway ensures the aggregation of data and such assures, that certain entities do their task but without getting individuals' data. An application of this privacy pattern would be the emission of the cars in Use Case 1. If we assume that each car can report about its emission, their emission values could be aggregated by a central authority. If all cars within an area form a group, the aggregation could give some indication about the impact on the air quality in that area.

4.2.7 Single point of contact

The Single Point of Contact orchestrates distributed service providers. In Use Case 1, a central authority would need to organise the different fog clusters along the interstate. The central authority could issue security tokens, authenticate local domain users and provide payment services, if the users make use of paid services.

5 Conclusion and future work

We applied seven privacy patterns to the IoT / Cloud Computing / Fog Computing architecture. By applying them to use cases from a smart vehicle scenario, we could demonstrate that they are applicable to real world scenarios. If used in the described manner, all of the privacy patterns can be used to improve users' privacy.

However, it is noteworthy that not all of the patterns can be applied in every case. In particular, the desire for certain properties of fog computing, e.g. a low latency, might prevent additional overhead caused by encryption or layers or redirection.

Additionally, with the lack of sufficient security protection causing IoT devices to be vulnerable to be hacked, broken or stolen [30], a general question arises. Is the data more secure if it is stored at the IoT nodes or at a central database of the cloud? To address this question one must make assumptions about possible and the most dangerous attackers in each case and in particular about the trustworthiness of the cloud and fog service providers. A general guideline is that the cloud and fog computing nodes will be more secure than the IoT nodes, so it will be less likely that they will be successfully attacked. On the other hand, the fog and cloud computing nodes are run by a third party with its own interests. Therefore, the question arises how trustworthy this party is.

In the same manner, it is not always clear, if users are able to control their data more easily if it is stored closer to them, but distributed (fog nodes) or if it is stored further away, but therefore centralised (cloud node).

We appreciate further research on the security and privacy relating to the storage of the data, the application of further privacy patterns to the IoT / Cloud Computing / Fog Computing architecture and the analysis of further examples, in particular if there is a trade-off between performance and privacy.

References

1. Evans D (2011) The Internet of Things How the Next Evolution of the Internet Is Changing Everything. Online White Paper. Available from: https://www.cisco.com/c/dam/en_us/about/ac79/docs/innov/IoT_IBSG_0411FINAL.pdf
2. Botta A, de Donato W, Persico V, Pescapé A (2016) Integration of Cloud computing and Internet of Things: A survey, Future Generation Computer Systems, Volume 56, p. 684–700, ISSN 0167-739X, Available from. https://doi.org/10.1016/j.future.2015.09.021
3. Thien AT, Colomo-Palacios R (2016) A Systematic Literature Review of Fog Computing. Paper presented at NOKOBIT 2016, Bergen, NOKOBIT, vol. 24, no. 1, Bibsys Open Journal Systems, ISSN 1894–7719
4. Kowatsch T, Maass TW (2012) *Privacy Concerns and Acceptance of IoT Services.* In: The Internet of Things 2012: New Horizons. Halifax, UK : IERC - Internet of Things European Research Cluster, S. 176–187
5. Fowler B (2017) Gifts That Snoop? The Internet of Things Is Wrapped in Privacy Concerns, Consumer Reports. Available from: https://www.consumerreports.org/internet-of-things/gifts-that-snoop-internet-of-things-privacy-concerns/
6. Hill K, Mattu S (2018) The House That Spied on Me, Gizmodo. Available from: https://gizmodo.com/the-house-that-spied-on-me-1822429852
7. Adams M (2017) Big Data and Individual Privacy in the Age of the Internet of Things. Technology Innovation Management Review 7(4):12–24
8. Papageorgiou A, Strigkos M, Politou E, Alepis E, Solanas A, Patsakis C (2018) Security and Privacy Analysis of Mobile Health Applications: The Alarming State of Practice. IEEE Access 6:9390–9403
9. Weinberg BD, Milne GR, Andonova YG, Hajjat FM (2015) Internet of Things: Convenience vs. privacy and secrecy. Business Horizons 58(6):615–624
10. Kristen L (2016) Walker: Surrendering information through the looking glass: Transparency, trust, and protection. J Public Policy Mark 35(1):144–158
11. Milne GR, Culnan MJ (2004) Strategies for reducing online privacy risks: Why consumers read (or don't read) online privacy notices. J Interact Mark 18(3):15–29
12. Milne GR, Culnan MJ, Greene H (2006) A longitudinal assessment of online privacy notice readability. J Public Policy Mark 25(2):238–249
13. Paul N, Tesfay W, Kipker D-K, Stelter M, Pape S (2018) Assessing Privacy Policies of Internet of Things Services. In ICT Systems Security and Privacy Protection - 33rd IFIP TC 11 International Conference, SEC 2018, Poznan
14. Iorga M, Goren N, Feldman L, Barton R, Martin M, Mahmoudi C (2018) Fog Computing Conceptual Model, NIST Special Publication 500–325, https://doi.org/10.6028/NIST.SP.500-325 available from: https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.500-325.pdf
15. Yousefpour A, Ishigaki G, Jue JP (2017) Fog Computing: Towards Minimizing Delay in the Internet of Things. 2017 IEEE International Conference on Edge Computing (EDGE), Honolulu, pp. 17–24
16. Bonomi F, Milito R, Zhu J, Addepalli S (2012) Fog computing and its role in the internet of things. In Proceedings of the first edition of the MCC workshop on Mobile cloud computing, p. 13–16. ACM
17. Bierzynski K, Escobar A, Eberl M (2017) Cloud, fog and edge: Cooperation for the future? FMEC: 62–67
18. Sathish Kumar J, Patel DR (2014) A survey on internet of things: Security and privacy issues. International Journal of Computer Applications 90.11
19. Martinez-Balleste A, Perez-Martinez PA, Solanas A (2013) The pursuit of citizens' privacy: a privacy-aware smart city is possible. IEEE Commun Mag 51(6):136–141
20. Dinev T, Hart P (2006) An Extended Privacy Calculus Model for E-Commerce Transactions. Inf Syst Res 17(1):61–80
21. Fred D (1989) Davis: Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Q 13(3):319–339
22. Kozlov D, Veijalainen J, Ali Y (2012) Security and privacy threats in IoT architectures. In: *Proceedings of the 7th International Conference on Body Area Networks* (BodyNets '12). ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), ICST, Brussels, 256–262
23. Lee K, Kim D, Ha D, Rajput U, Oh H (2015) On security and privacy issues of fog computing supported Internet of Things environment. In: Network of the Future (NOF), 2015 6th International Conference on the, pp. 1–3. IEEE
24. Stojmenovic I, Wen S (2014) The Fog Computing Paradigm: Scenarios and Security Issues. FedCSIS 1–8
25. Stojmenovic I, Wen S, Huang X, Luan H (2016) An overview of Fog computing and its security issues. Concurrency and Computation: Practice and Experience 28(10):2991–3005
26. Lu R, Liang X, Li X, Lin X, Shen X (2012) Eppa: An efficient and privacy-preserving aggregation scheme for secure smart grid communications. Parallel and Distributed Systems, IEEE Transactions 23(9):1621–1631
27. Ni J, Zhang K, Lin X, Shen X (2017) Securing fog computing for internet of things applications: Challenges and solutions. IEEE Communications Surveys & Tutorials
28. Tayeb S, Latifi S, Kim Y (2017) A survey on IoT communication and computation frameworks: An industrial perspective. In: Computing and Communication Workshop and Conference (CCWC), 2017 IEEE 7th Annual, pp. 1–6. IEEE
29. Sadeghi A-R, Wachsmann C, Waidner M (2015) Security and privacy challenges in industrial Internet of Things. Design Automation Conf. (DAC), 2015 52nd ACM/EDAC/IEEE, pp. 1–12
30. Yi S, Qin Z, Li Q (2015) Security and privacy issues of fog computing: a survey. In Wireless Algorithms, Systems, and Applications 2015 (pp. 685–695), Springer International Publishing. Available from http://link.springer.com/chapter/10.1007/978-3-319-21837-3_67
31. Rahman LF, Ozcelebi T, Lukkien JJ (2016) Choosing your IoT programming framework: Architectural aspects. In: Future Internet of Things and Cloud (FiCloud), 2016 IEEE 4th International Conference on, pp. 293–300. IEEE
32. Graf C, Wolkerstorfer P, Geven A, Tscheligi M (2010) A pattern collection for privacy enhancing technology. In: The 2nd Int. Conf. on Pervasive Patterns and Applications (PATTERNS 2010), pp. 21–26
33. Yoder J, Baraclow J (1997) Architectural Patterns for Enabling Application Security. Pattern Languages of Programs
34. International Standards Organisation (1999) Common criteria for information technology security evaluation. http://www.commoncriteriaportal.org/public/consumer/index.php?menu=2
35. Schumacher M (2002) Security Patterns and Security Standards - With Selected Security Patterns for Anonymity and Privacy. European Conference on Pattern Languages of Programs (EuroPLoP)
36. Privacy Patterns Website. https://privacypatterns.org
37. Schümmer T (2004) The Public Privacy – Patterns for Filtering Personal Information in Collaborative Systems. CHI

38. Romanosky S, Acquisti A, Hong J, Cranor LF, Friedman B (2006) Privacy patterns for online interactions. In: Proceedings of the 2006 conference on Pattern languages of programs. ACM, p. 12
39. Doty N, Gupta M (2013) Privacy design patterns and anti-patterns. In: Trustbusters Workshop at the Symposium on Usable Privacy and Security
40. Privacy Design Pattern Library Website. http://www.legaltechdesign.com/communication-design/legal-design-pattern-libraries/privacy-design-pattern-library/
41. Paillier P (1999) Public-key cryptosystems based on composite degree residuosity classes. In International Conference on the Theory and Applications of Cryptographic Techniques, pp. 223–238. Springer, Berlin, Heidelberg
42. Shamir A (1979) How to share a secret. Commun ACM 22(11): 612–613
43. Dierks T (2008) The transport layer security (TLS) protocol version 1.2, RFC 5246
44. Okay FY, Ozdemir S (2018) A secure data aggregation protocol for fog computing based smart grids. 2018 IEEE 12th International Conference on Compatibility, Power Electronics and Power Engineering (CPE-POWERENG 2018), Doha, pp. 1–6
45. Rannenberg K (2016) Opportunities and Risks Associated with Collecting and Making Usable Additional Data. Autonomous Driving. Springer, Berlin, Heidelberg, 497–517
46. SAE (2014) Taxonomy and definitions for terms related to on-road-motor vehicle automated deriving systems, J3016, SAE International Standard

C.7 Why Do People Pay for Privacy?

David Harborth, Xinyuan Cai, and Sebastian Pape. Why do people pay for privacy? In *ICT Systems Security and Privacy Protection - 34th IFIP TC 11 International Conference, SEC 2019, Lisbon, Portugal, June 25-27, 2019, Proceedings*, pages 253–267, 06 2019. doi: 10.1007/978-3-030-22312-0_18. URL https://doi.org/10.1007/978-3-030-22312-0_18

Why Do People Pay for Privacy-Enhancing Technologies? The Case of Tor and JonDonym

David Harborth(✉), Xinyuan Cai, and Sebastian Pape

Chair of Mobile Business and Multilateral Security,
Goethe University Frankfurt, Frankfurt am Main, Germany
david.harborth@m-chair.de

Abstract. Today's environment of data-driven business models relies heavily on collecting as much personal data as possible. One way to prevent this extensive collection, is to use privacy-enhancing technologies (PETs). However, until now, PETs did not succeed in larger consumer markets. In addition, there is a lot of research determining the technical properties of PETs, i.e. for Tor, but the use behavior of the users and, especially, their attitude towards spending money for such services is rarely considered. Yet, determining factors which lead to an increased willingness to pay (WTP) for privacy is an important step to establish economically sustainable PETs. We argue that the lack of WTP for privacy is one of the most important reasons for the non-existence of large players engaging in the offering of a PET. The relative success of services like Tor corroborates this claim since this is a service without any monetary costs attached. Thus, we empirically investigate the drivers of active users' WTP of a commercial PET - JonDonym - and compare them with the respective results for a donation-based service - Tor. Furthermore, we provide recommendations for the design of tariff schemes for commercial PETs.

Keywords: Privacy · Privacy-enhancing technologies · Pricing · Willingness to pay · Tor · JonDonym

1 Introduction

Perry Barlow states: "The internet is the most liberating tool for humanity ever invented, and also the best for surveillance. It's not one or the other. It's both" [1]. One of the reasons for surveilling users is a rising economic interest in the internet [2]. However, users who have privacy concerns and feel a strong need to protect their privacy are not helpless, they can make use of privacy-enhancing technologies (PETs). PETs allow users to improve their privacy by eliminating or minimizing personal data disclosure to prevent unnecessary or unwanted processing of personal data [3]. Examples of PETs include services which allow anonymous communication, such as Tor [4] or JonDonym [5]. There has been lots of research on Tor and JonDonym [6, 7], but the large majority of it is of technical nature and does not consider the user. However, the number of users if crucial for this kind of services. Besides the economic point of view which suggests that more users allow a more cost-efficient way to run those services, the quality of the offered service is depending on the number of users

Published by Springer Nature Switzerland AG 2019
G. Dhillon et al. (Eds.): SEC 2019, IFIP AICT 562, pp. 253–267, 2019.
https://doi.org/10.1007/978-3-030-22312-0_18

since an increasing number of (active) users also increases the anonymity set. The anonymity set is the set of all possible subjects who might cause an action [8], thus a larger anonymity set may make it more difficult for an attacker to identify the sender or receiver of a message.

In the end, the sustainability of a service not only depends on the number of active users but also on a company or organization with the intention of running the service. One intention certainly is a well working business model. As a consequence, it is crucial to not only learn about the users' intention to use a PET, but also to understand the users' willingness to pay (WTP) for a service. Determining factors to understand the users' WTP along with a suitable tariff structure is the key step to establish economically sustainable services for privacy. The current market for PET providers is rather small, some say the market even fails [9]. We argue that the lack of WTP for privacy is one of the most important reasons for the non-existence of large players engaging in the offering of a PET. Earlier research on WTP often works with hypothetical scenarios (e.g. with conjoint-analyses) and concludes that users are not willing to pay for their privacy [10, 11]. We tackle the issue based on actual user experiences and behaviors and enhance the past research by analyzing two existing PETs with active users, with some of them already paying or donating for the service. Tor and JonDonym are comparable with respect to their functionality and partially with respect to the users' perceptions about them. However, they differ in their business model and organizational structure. Therefore, we investigate the two research questions:

RQ1: Which factors influence the willingness to pay for PETs?
RQ2: What are preferred tariff options of active users of a commercial PET?

The remainder of the paper is structured as follows: Sect. 2 briefly introduces the anonymization services Tor and JonDonym and lists related work on PETs and users' willingness to pay. In Sect. 3, we present the research hypotheses and describe the questionnaire and the data collection process. We present the results of our empirical research in Sect. 4 and discuss the results and conclude the paper in Sect. 5.

2 Theoretical Background and Related Work

Privacy-Enhancing Technologies (PETs) is an umbrella term for different privacy protecting technologies. Borking and Raab define PETs as "a coherent system of ICT measures that protects privacy […] by eliminating or reducing personal data or by preventing unnecessary and/or undesired processing of personal data; all without losing the functionality of the data system" [12]. In the following sections, we describe Tor and JonDonym as well as related work with respect to WTP for privacy.

2.1 Tor and JonDonym

Tor and JonDonym are low latency anonymity services which redirect packets in a certain way in order to hide metadata (the sender's/receiver's internet protocol (ip) address) from passive network observers. Low latency anonymity services can be used for interactive services such as messengers. Due to network overheads this still

leads to increased latency which was evaluated by Fabian et al. [13] who found associated usability issues when using Tor. Technically, Tor – the onion router – is an overlay network where the users' traffic is encrypted and directed over several different servers (relays). The chosen traffic routes should be difficult for an adversary to observe, which means that unpredictable routes through the Tor network are chosen. The relays where the traffic leaves the tor network are called "exit nodes" and for an external service the traffic seems to originate from those. JonDonym is based on user selectable mix cascades, with two or three mix servers in one cascade. For mix networks route unpredictability is not important so within one cascade always the same sequence of mix servers is used. Thus, for an external service the traffic seems to originate from the last mix server in the cascade. As a consequence, other usability issues may arise when websites face some abusive traffic from the anonymity services [14] and decide to restrict users from the same origin. Restrictions range from outright rejection to limiting the users' access to a subset of the services functionality or imposing hurdles such as CAPTCHA-solving [15]. For the user it appears that the website is not function properly. Tor offers an adapted browser including the Tor client for using the Tor network, the "Tor Browser". Similarly, the "JonDoBrowser" includes the JonDo client for using the JonDonym network. Although technically different, JonDonym and Tor are highly comparable with respect to the general technical structure and the use cases. However, the entities who operate the PETs are different. Tor is operated by a non-profit organization with thousands of voluntarily operated servers (relays) over which the encrypted traffic is directed. Tor is free to use with the option that users can donate to the Tor project. The actual number of users is estimated with approximately 2,000,000 active users [4]. JonDonym is run by a commercial company. The mix servers used to build different mix cascades are operated by independent and non-interrelated organizations or private individuals who all publish their identity. The service is available for free with several limitations, like the maximum download speed. In addition, there are different premium rates without these limitations that differ with regard to duration and included data volume. Thus, JonDonym offers several different tariffs and is not based on donations. The actual number of users is not predictable since the service does not keep track of this.

From a research perspective, there are some papers about JonDonym, e.g. a user study on user characteristics of privacy services [16]. Yet, the majority of work is about Tor. Most of the work is technical [6], e.g. on improvements such as relieved network congestion, improved router selection, enhanced scalability or reduced communication/computational cost of circuit construction [17]. There is also lots of work about the security respectively anonymity properties [18, 19] and traffic correlation [20].

2.2 Related Work

Previous non-technical work on PETs mainly considers usability studies and does not primarily focus on WTP. For example, Lee et al. [21] assess the usability of the Tor Launcher and propose recommendations to overcome the found usability issues. Further research suggests zero-effort privacy [22, 23] by improving the usability of the service. In quantitative studies, we already investigated privacy concerns and trust on

JonDonym [24] and Tor [25, 26] based on Internet users' information privacy concerns (IUIPC) [27] and could extent the causal model by "trust in the service" which plays a crucial role for the two PETs. Some experiments suggest that users are not willing to pay for their privacy [10, 11]. In contrast to these experiments, we surveyed actual users – some of them already paying or donating for the service. Grossklags find contradicting behavior of users when it comes to WTP to protect information and "willingness to accept" compensation for revealing information [28]. Further work covers selling personal data [29, 30] e.g. on data markets [31] or experiments on the value of privacy [32]. Some work tries to explain the privacy paradox with economic models [33] or discusses the right of the users to know the value of their data [34]. However, all of these are focused on the value of certain data or privacy and not on the users' WTP for privacy. Cranor et al. investigate how actual users use their privacy preferences tool [35]. Spiekermann investigate the traits and views of actual users of the predecessor of JonDonym, AN.ON/JAP, a free anonymity service [16]. However, since the tools were free, none of them investigated the users' WTP. Following a more high-level view, some research addresses the markets for PETs. Federrath claims that there is a market for PETs but they have to consider law enforcement functionality [36]. Rossnagel analyzes PET markets based on diffusion of innovations theory about anonymity services [9] and concludes a market failure. Schomakers et al. do a cluster analysis of users and find three groups with different attitudes towards privacy and argue that each of the groups need distinct tools [37]. In the same line, further research concludes that one should focus on specific subgroups for the adoption of Tor [38]. Following a market perspective, Boehme et al. analyze the condition under which it is profitable for sellers in e-commerce environments to support PETs, assuming that without PETs they could increase their profit with price discrimination [39].

3 Methodology

In this section we present the research hypotheses, the questionnaire and the data collection process. The demographic questions were not mandatory to fill out. This was done on purpose since we assumed that most of the participants are highly sensitive with respect to their personal data and could potentially react to mandatory demographic questions by terminating the survey. Consequently, the demographics are incomplete to a large extent. Therefore, we had to resign from a discussion of the demographics in our research context.

The statistical analysis of the research data is conducted with the open-source software R. First of all, we focus solely on JonDonym and compare the differences of average preferences for alternative tariff schemes. Thereby, we differentiate between participants stating to use JonDonym in the free of charge option those stating to use it with one of the available premium tariffs. Due to non-normality of the data, we use the non-parametric test Wilcoxon rank sum test to determine whether preferences for newly designed tariffs differ from each other among different types of users. We designed these new tariffs in collaboration with the chief executive of the JonDos

GmbH in order to provide realistic pricing schemes which are economically viable and sustainable for the company. We used the paired Wilcoxon test to determine whether users' preferences for one tariff are statistically significantly different from the other tariffs. The Wilcoxon rank sum test is also called Mann-Whitney-U-Test. It is a non-parametric test of the null hypothesis that the mean of one sample will be different from the mean from a second sample. The paired Wilcoxon test is also called the Wilcoxon signed-rank test which is a similar nonparametric test used for dependent samples [40, 41]. In order to illustrate the difference in preferences among two types of users, i.e. free users and premium users, we use boxplots to visualize the descriptive statistics of the two samples [42]. A boxplot is a method for graphically depicting groups of numerical data through their quartiles. Boxplots are non-parametric. They display variation in samples of a statistical population without making any assumptions of the underlying statistical distribution. The upper line of the box is the first quartile, the band inside the box is the second quartile (the median) and the bottom line of the box is the third quartile.

3.1 Research Model and Hypotheses for the Logistic Regression Model

As a last step, we conduct a logistic regression to find out which factors influence users' willingness to pay for privacy (in our case willingness to pay for JonDonym and willingness to donate to Tor). We used the logistics regression to build the model because our dependent variable is a binary variable. A linear regression is not an appropriate model here due to the violation of the assumption that the dependent variable (WTP) is continuous, with errors which are normally distributed [43]. The probit regression is also not suitable because it assumes that our dependent variable is not normally distributed. Willingness to pay for JonDonym is defined as the binary classification of JonDonym users' actual behavior.

$$willingness\,to\,pay = \begin{cases} 0, if\ the\ respondent\ uses\ a\ free\ tariff \\ 1, if\ the\ respondent\ uses\ a\ premium\ tariff \end{cases} \quad (1)$$

Accordingly, willingness to donate is defined as the binary classification of Tor users' actual behavior.

$$willingness\,to\,donate = \begin{cases} 0, if\ the\ respondent\ has\ never\ donated \\ 1, if\ the\ respondent\ has\ donated \end{cases} \quad (2)$$

The independent variables are risk propensity (RP), frequency of improper invasion of privacy (VIC), trusting beliefs in online companies (TRUST), trusting beliefs in JonDonym ($TRUST_{PET}$) and knowing of Tor/JonDonym (TOR/JD) or not. Thus, our research model is as follows:

$$WTP/WTD_i = \beta_0 + \beta_1 RP_i + \beta_2 VIC_i + \beta_3 TRUST_i + \beta_4 TRUST_{PET,i} + \beta_5 TOR/JD_i + \varepsilon_i \quad (3)$$

Risk propensity measures the risk aversion of the individual, i.e. the higher the measure, the more risk-averse the individual [44]. Literature finds that a risk aversion can act as a driver to protect an individual's privacy [45]. Thus, we hypothesize:

H1: Risk propensity (RP) has a positive effect on the likelihood of paying or donating for PETs.

Privacy victim (VIC) measures how often individuals experienced a perceived improper invasion in their privacy [27]. Results of past research dealing with perceived bad experiences with privacy indicate that such experiences can cause individuals to protect their privacy to a larger extent [46]. Thus, we hypothesize:

H2: The more frequent users felt that they were a victim of an improper breach of their privacy, the more likely they are to pay or donate for PETs.

The construct *trust in online companies* assesses individuals trust in online companies with respect to handling their personal data [27]. Results in the literature suggest that a higher trust in online companies has a positive effect on the willingness to disclose personal information. Following this finding, we argue that users who have a higher level of trust in online companies, are less likely to spend money for protecting their privacy. Therefore, we hypothesize:

H3: The more users trust online companies with handling their personal data, the less likely they are to pay or donate for PETs.

Trust in JonDonym/Tor is adapted from Pavlou [47]. Trust can refer to the technology (in our case PETs (Tor and JonDonym)) itself as well as to the service provider. Since the non-profit organization of Tor evolved around the service itself [4], it is rather difficult for users to distinguish which label refers to the technology itself and which refers to the organization. The same holds for JonDonym since JonDonym is the only main service offered by the commercial company JonDos. Therefore, we argue that it is rather difficult for users to distinguish which label refers to the technology itself and which refers to the company. Thus, we decided to ask for trust in the PET (Tor and JonDonym, respectively), assuming that the difference to ask for trust in the organization/company is negligible. Literature shows that trust in services enables positive attitudes towards interacting with these services [24–26, 47]. In line with these results, we argue that a higher level of trust in the PET increases the likelihood to spend money for it. Thus, we hypothesize:

H4: The more users trust the PET, the more likely they are to pay or donate for it.

Lastly, we included a question about whether users of Tor/JonDonym know JonDonym/Tor. We included this question due to previous findings about a substituting effect of Tor with regard to the WTP for JonDonym [48]. Users of JonDonym partially stated that they would only spend money for a premium tariff, if Tor was not existent. Thus, we wanted to include this factor as a control variable in our analysis and hypothesize:

H5: The likelihood of JonDonym users to pay for a premium tariff decreases, if they are aware of Tor (we do not expect a similar effect for Tor users).

3.2 Data Collection

We conducted the studies with German and English-speaking users of Tor and JonDonym. For each service, we administered two questionnaires. Partially, items for the German questionnaire had to be translated since some constructs are adapted from the English literature. To ensure content validity of the translation, we followed a rigorous translation process. First, we translated the English questionnaire into German with the help of a certified translator (translators are standardized following the DIN EN 15038 norm). The German version was then given to a second independent certified translator who retranslated the questionnaire to English. This step was done to ensure the equivalence of the translation. Third, a group of five academic colleagues checked the two English versions with regard to this equivalence. All items were found to be equivalent [49]. The items for all analyses can be found in the appendix.

We installed the surveys on a university server and managed it with the survey software LimeSurvey (version 2.72.6) [50]. For Tor, we distributed the links to the English and German version over multiple channels on the internet. An overview of every distribution channel can be found in an earlier paper based on the same dataset [26]. In sum, 314 participants started the questionnaire (245 English version, 40 English version posted in hidden service forums, 29 German version). Of those 314 approached participants, 135 (105 English version, 13 English version posted in hidden service forums, 17 German version) filled out the questionnaires completely. After deleting all participants who answered a test question in the middle of the survey incorrectly, 124 usable data sets remained for the following analysis. For JonDonym, we distributed the links to the English and German version with the beta version of the JonDonym browser and published them on the official JonDonym homepage. In sum, 416 participants started the questionnaire (173 English version, 243 German version). Of those 416 approached participants, 141 (53 English version, 88 German version) remained after deleting unfinished sets and all participants who answered a test question incorrectly.

4 Results

We present the results of our empirical analyses in this section. In the first part, we discuss the analysis of the current tariff structures (JonDonym) and donation statistics (Tor). Furthermore, we assess preferences of JonDonym users regarding new alternative tariff schemes. In the second part, we show the results of the logistic regression model with the factors influencing the willingness to pay (JonDonym)/to donate (Tor).

4.1 Tariff Analysis for JonDonym

Among the 141 JonDonym users in of our survey, 85 users use a free tariff. 56 users are using JonDonym with a paid tariff. Among the 124 Tor users of our survey, 93 of them have never donated to Tor. Among donating users, the amounts of donation are arbitrary. The payment structure of JonDonym and descriptive statistics for the donations to Tor are shown in Table 1. It can be seen that roughly 1/3 of the participants spend

money for JonDonym (25%) and Tor (39.72%). To analyze potential tariff optimizations for JonDonym, we asked about users' preferences for three general tariff structures, namely a high-data-volume tariff (TP1), a low-price tariff (TP2) and a low-anonymity tariff (TP3). In addition, we designed five new tariffs. TRN4 is the tariff with the lowest data volume per month and TRN5 is the tariff with highest data volume per month. The specific wording of the tariff options can be found in the appendix.

Table 1. Tariff and donation statistics of JonDonym and Tor users

JonDonym		Tor	
Tariff option	N = 141	Tariff option	N = 124
Free of charge option	85	No donation	93
Volume-M (1500 MB/12 months 10€)	28	Donation	31
Volume-L (5000 MB/24 months 30€)	19	Min. donation	0.00
Flat-M (monthly 2 GB/6 months/50€)	5	Median donation	100.00
Flat-L (monthly 5 GB/6 months/100€)	4	Mean donation	301.40
Volume-S (650 MB/6 months 5€)	0	Max. donation	4500.00

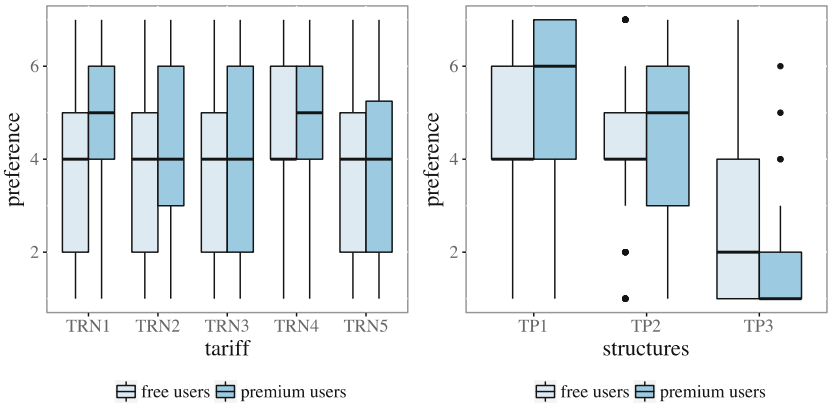


Fig. 1. Users' preference for alternative tariff structures (left side) and users' preferences for tariff structures (right side), free users = 85, premium users = 56

Figure 1 shows the boxplots for the preferences for the five new tariff options (TRN) differentiated between free and premium users as well as three alternative tariff structures (TP). The median preferences of free users for the five tariffs are neutral (preference = 4). However, the mean preference of free users for TRN4 is slightly higher compared to the other options. In comparison, premium users have a higher preference for TRN1 and TRN4. In a next step, we analyze whether the differences illustrated with the boxplots between options for the different groups (full sample,

premium users, free users) are statistically significant (Table 2). Our results indicate that the whole sample of users shows the highest preference for TRN4 and the second highest preference for TRN1. The remaining tariffs, i.e. TRN2, TRN3 and TRN5 are favored least of all. However, this contradicts with the conclusion that the total users show the highest preference for TP1. Thus, it makes sense to split the sample and look at free and premium users. Premium users show the highest preference for TRN1 and TRN4, the second highest preference for TRN2 and TRN3, and the least preference for TRN5. Thus, they show a higher preference for 100 GB tariffs. This is in line with the conclusion that premium users have the highest preference for TP1. Free users show a neutral preference for all five tariffs except for TRN4 (slightly higher).

Table 2. Paired Wilcoxon tests for the five new tariffs and three tariff structures

New tariffs/structures		*reject* H_0*: X = Y* N = 141 Total users	*reject* H_0*: X = Y* N = 56 Premium users	*reject* H_0*: X = Y* N = 85 Free users
X	Y			
TRN1	TRN2	Yes*	Yes**	No
TRN1	TRN3	Yes**	Yes*	No
TRN1	TRN4	Yes*	No	Yes***
TRN1	TRN5	Yes*	Yes**	No
TRN2	TRN3	No	No	No
TRN2	TRN4	Yes***	No	No
TRN2	TRN5	No	No	No
TRN3	TRN4	Yes***	Yes*	Yes**
TRN3	TRN5	No	No	No
TP1	TP2	Yes*	Yes***	No
TP1	TP3	Yes ***	Yes***	Yes***
TP2	TP3	Yes ***	Yes***	Yes***

**significance level of paired Wilcoxon test:* *p < 0.05, **p < 0.01, ***p < 0.001

Table 2 also presents the results for the differences in preferences for the tariff structures (TP). The results indicate that the 141 users have a higher preference for a high-data-volume tariff compared to a low-price tariff (TP1 vs. TP2). The results are similar for the sub-group of premium users. They have the same preference order as the whole sample of users. However, free users have the same preference for TP1 and TP2.

4.2 Factors Influencing Willingness to Pay for Privacy

Before analyzing the results in detail, we have to assess whether the independent variables correlate with each other (multicollinearity), since this would negatively impact the validity of our model. We test for multicollinearity by calculating the variance inflation factor (VIF) for all independent variables. None of the variables has a VIF larger than 1.7, indicating that multicollinearity is not an issue for our sample.

The results of the logistic regression model can be seen in Table 3. We highlighted statistically significant results in bold face. For JonDonym, RP and $TRUST_{PET}$ are the only statistically significant independent variables in the model. Surprisingly, RP has a negative coefficient, indicating that more risk-averse users are less likely to choose a premium tariff for JonDonym. This empirical result is in contrast to hypothesis 1, thus we cannot confirm this hypothesis derived from results of the literature and the associated rationale. Reasons for this contradictory result can be manifold. For example, there might be unobservable variables not included in the model which impact the relationship between RP and WTP. Hypotheses 2, 3 and 5 cannot be confirmed as well due to insignificant coefficients. In contrast to this, hypothesis 4 can be confirmed. Given the average marginal effect (avg. marg. effect), our result indicates that a one unit increase in trust in JonDonym increases the likelihood of choosing a premium tariff by 12.17%. This result is statistically significant at the 0.1% level. Hypothesis 4 can also be confirmed for the logistic regression model for Tor users with a slightly larger average marginal effect size of 12.45%. The variable VIC is statistically significant at the 1% level with a marginal effect of 5.33%. This indicates that bad experiences with privacy breaches lead to a higher probability of donating money to Tor, and thereby, supporting the Tor project financially. No other hypotheses can be confirmed for Tor.

Table 3. Results of the logistic regression model

	WTP for JonDonym		WTD for Tor		Difference
	Coef.	Avg. marg. effects	Coef.	Avg. marg. effects	Avg. marg. effects
(Intercept)	−0.0376	−0.0081	6.1455***	−0.9768	0.9687
RP	**−0.4967****	**−0.1067**	−0.1492	−0.0237	−0.083
VIC	−0.0397	−0.0085	**0.3352****	**0.0533**	−0.0618
TRUST	−0.0868	−0.0187	−0.1222	−0.0194	0.0007
$TRUST_{PET}$	**0.5661*****	**0.1217**	**0.7835*****	**0.1245**	−0.0028
TOR/JD	−0.5792	−0.1245	0.488	0.0776	−0.2021

*p < 0.05, **p < 0.01, ***p < 0.001

5 Discussion and Conclusion

With respect to research question 1, our results show that PET providers should focus on building a strong reputation since trust in the PET is the strongest factor influencing the probability of spending money for privacy for both, JonDonym and Tor. In addition, we can observe that Tor users are more likely to donate for the service if they were a victim of a privacy breach or violation in their past.

Our second research question is about an optimized design of tariff options for users of commercial PETs based on the case of JonDonym. Here, we can see that the results differ when looking at different groups of users, which is in line with former research [37].

Users who use JonDonym with the free option, are indifferent with respect to the newly introduced tariffs as well as the general tariff structures (high volume vs. low price vs. low anonymity). However, some of them tend to prefer the tariff option with the lowest price with an included high-speed volume of 40 GB the most. Thus, free users would prefer the cheapest tariff, if they were to decide for paying at all. Practically, this implies that commercial PET providers should try to offer options with a relatively low monetary barrier to convert as many free users as possible into paying ones. The already paying users prefer high-volume tariffs over the other options.

Limitations of this study are the following. First, our sample only includes a relatively small number of active users of both PETs. This sample size is sufficient for the sake of our statistical analyses. However, the results about the current payment and donation numbers provide only a rough idea about the actual distribution. In addition, it is very difficult to gather data of actual users of PETs since it is a comparable small population that we could survey. It is also relevant to mention that we did not offer any financial rewards for the participation. A second limitation concerns possible self-report biases (e.g. social desirability). We addressed this issue by gathering the data fully anonymized. Third, mixing results of the German and English questionnaire could be a source of errors. On the one hand, this procedure was necessary to achieve the minimum sample size. On the other hand, we followed a very thorough translation procedure to ensure the highest level of equivalence as possible. Thus, we argue that this limitation did not affect the results to a large extent. However, we cannot rule out that there are unobserved effects on the results due to running the survey in more than one country at all. Lastly, demographic questions were not mandatory to fill out due to our assumption that these types of individuals who use Tor or JonDonym are highly cautious with respect to their privacy. Thus, we decided to go for a larger sample size considering that we might have lost participants otherwise (if demographics had to be filled out mandatorily). However, we must acknowledge that demographic variables might be relevant confounders in the regression model explaining the WTP of PET users.

Future work should aim to determine the relation between paying users and the groups Schomakers et al. [37] identified. In addition, researchers can build on our results by implementing such tariff options for commercial PET services in practice and investigate whether users are more prone to spend money for their privacy protection. Furthermore, it is relevant for commercial PET providers to differentiate themselves against free competitors as Tor in our example. This can be done by providing a higher level of usability in terms of ease of use, performance and compatibility with other applications [25, 48]. If commercial PET providers cannot create a unique selling point (USP) compared to free services, it is very unlikely that they establish a successful monetarization strategy in the market. Therefore, it is necessary to investigate how a USP for a commercial PET provider can look like and assess it in the field with active users of existing PETs as well as non-users.

Appendix - Questionnaire

A. Constructs and Questions for both PETs

Risk Propensity (RP)
1. I would rather be safe than sorry.
2. I am cautious in trying new/ different products.
3. I avoid risky things.

Trust in the PET (JonDonym / Tor) ($TRUST_{PET}$)
1. JonDonym / Tor is trustworthy.
2. JonDonym / Tor keeps promises and commitments.
3. I trust JonDonym / Tor because they keep my best interests in mind.

Trust in Online Companies (TRUST)

1. Online companies are trustworthy in handling information.
2. Online companies tell the truth and fulfill promises related to information provided by me.
3. I trust that online companies would keep my best interests in mind when dealing with information.
4. Online companies are in general predictable and consistent regarding the usage of information.
5. Online companies are always honest with customers when it comes to using the provided information.

Privacy Victim (VIC)
How frequently have you personally been the victim of what you felt was an improper invasion of privacy? (7-point frequency scale from "Never" to "Very frequently")
Knowledge about Tor (TOR)/JonDonym (JD)
Do you know the anonymization service Tor/JonDonym? (Yes/No)

B. Specific Questions for JonDonym
Current Tariff - Please choose your current tariff of JonDonym.

1. Free of charge option
2. Flat-M (monthly 2GB / 6 months / 50€)
3. Flat-L (monthly 5GB / 6 months / 100€)
4. Volume-S (650 MB / 6 months 5€)
5. Volume-M (1500 MB / 12 months 10€)
6. Volume-L (5000 MB / 24 months 30€)

Tariff Preference (TP)

1. I would use JD regularly with a data volume ten times higher than before (at the same price).
2. If the price decreased by half, I would use JonDonym regularly.
3. I would perceive a service with a lower anonymization level for half the price more attractive than JonDonym.

Tariff New (TRN)

1. Monthly 100 GB with a duration of 12 months for 100€ (total price)
2. Monthly 100 GB with a duration of 3 months for 30€ (total price)
3. Monthly 100 GB with a duration of 12 months for 10€ per month

4. Monthly 40 GB with a duration of 3 months for 5€ per month
5. Monthly 200 GB with a duration of 12 months for 15€ per month

C. Specific Questions for Tor

Donation to Tor

Did you ever donate money to the Tor project? (Yes/No)

Donation Amount

How much money did you donate to the Tor project? (open field with number only)

If not stated otherwise, constructs are measured based on a 7-point Likert scale ranging from strongly disagree to strongly agree.

References

1. Ball, J.: Hacktivists in the frontline battle for the internet. https://www.theguardian.com/technology/2012/apr/20/hacktivists-battle-internet
2. Bédard, M.: The Underestimated Economic Benefits of the Internet. In: Regulation Series. The Montreal Economic Institute (2016)
3. van Blarkom, G.W., Borking, J.J., Olk, J.G.E.: PET: Handbook of Privacy and Privacy-Enhancing Technologies (2003)
4. The Tor Project: Tor. https://www.torproject.org
5. JonDos Gmbh: Official Homepage of JonDonym. https://www.anonym-surfen.de
6. Saleh, S., Qadir, J., Ilyas, M.U.: Shedding light on the dark corners of the internet: A survey of tor research. J. Netw. Comput. Appl. **114**, 1–28 (2018)
7. Montieri, A., Ciuonzo, D., Aceto, G., Pescapé, A.: Anonymity services Tor, I2P, JonDonym: classifying in the dark. In: International Teletraffic Congress, pp. 81–89 (2017)
8. Pfitzmann, A., Hansen, M.: A terminology for talking about privacy by data minimization: anonymity, unlinkability, undetectability, unobservability, pseudonymity, and identity management, pp. 1–98. Tech. Univ. Dresden (2010)
9. Rossnagel, H.: The market failure of anonymity services. In: Samarati, P., Tunstall, M., Posegga, J., Markantonakis, K., Sauveron, D. (eds.) WISTP 2010. LNCS, vol. 6033, pp. 340–354. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-12368-9_28
10. Grossklags, J., Acquisti, A.: When 25 cents is too much: an experiment on willingness-to-sell and willingness-to-protect personal information. In: WEIS (2007)
11. Beresford, A.R., Kübler, D., Preibusch, S.: Unwillingness to pay for privacy: a field experiment. Econ. Lett. **117**, 25–27 (2012)
12. Borking, J.J., Raab, C.: Laws, PETs and other technologies for privacy protection. J. Inf. Law Technol. **1**, 1–14 (2001)
13. Fabian, B., Goertz, F., Kunz, S., Müller, S., Nitzsche, M.: Privately waiting – a usability analysis of the tor anonymity network. In: Nelson, M.L., Shaw, M.J., Strader, T.J. (eds.) AMCIS 2010. LNBIP, vol. 58, pp. 63–75. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-15141-5_6
14. Singh, R., et al.: Characterizing the nature and dynamics of Tor exit blocking. In: 26th USENIX Security Symposium (USENIX Security), Vancouver, BC, pp. 325–341 (2017)
15. Chirgwin, R.: CloudFlare shows Tor users the way out of CAPTCHA hell. https://www.theregister.co.uk/2016/10/05/cloudflare_tor/
16. Spiekermann, S.: The desire for privacy: insights into the views and nature of the early adopters of privacy services. Int. J. Technol. Hum. Interact. **1**, 74–83 (2005)

17. Alsabah, M., Goldberg, I.: Performance and security improvements for tor: a survey. ACM Comput. Surv. (CSUR) **49**(2), 1–36 (2016). Article no. 32
18. Koch, R., Golling, M., Rodosek, G.D.: How anonymous is the tor network? A long-term black-box investigation. Computer (Long. Beach. Calif.) **49**, 42–49 (2016)
19. Juarez, M., Elahi, T., Jansen, R., Diaz, C., Galvez, R., Wright, M.: Poster: fingerprinting hidden service circuits from a tor middle relay. In: Proceedings of IEEE S&P (2017)
20. Johnson, A., Wacek, C., Jansen, R., Sherr, M., Syverson, P.: Users get routed: traffic correlation on tor by realistic adversaries. In: ACM CCS, pp. 337–348 (2013)
21. Lee, L., Fifield, D., Malkin, N., Iyer, G., Egelman, S., Wagner, D.: A usability evaluation of tor launcher. In: Proceedings on Privacy Enhancing Technologies, pp. 90–109 (2017)
22. Herrmann, D., Lindemann, J., Zimmer, E., Federrath, H.: Anonymity online for everyone: what is missing for zero-effort privacy on the internet? In: Camenisch, J., Kesdoğan, D. (eds.) iNetSec 2015. LNCS, vol. 9591, pp. 82–94. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-39028-4_7
23. Harborth, D., et al.: Integrating privacy-enhancing technologies into the internet infrastructure. arXiv Prepr. arXiv:1711.07220 (2017)
24. Harborth, D., Pape, S.: JonDonym users' information privacy concerns. In: Janczewski, L.J., Kutyłowski, M. (eds.) SEC 2018. IAICT, vol. 529, pp. 170–184. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-99828-2_13
25. Harborth, D., Pape, S.: Examining technology use factors of privacy-enhancing technologies: the role of perceived anonymity and trust. In: Twenty-Fourth Americas Conference on Information Systems, New Orleans, USA (2018)
26. Harborth, D., Pape, S.: How privacy concerns and trust and risk beliefs influence users' intentions to use privacy-enhancing technologies - the case of tor. In: Hawaii International Conference on System Sciences Proceedings, Hawaii, US (2019)
27. Malhotra, N.K., Kim, S.S., Agarwal, J.: Internet users' information privacy concerns: the construct, the scale, and a causal model. Inf. Syst. Res. **15**, 336–355 (2004)
28. Grossklags, J.: Experimental economics and experimental computer science: a survey. In: Workshop on Experimental Computer Science, ExpCS 2007 (2007)
29. Acquisti, A.: The economics of personal data and the economics of privacy. In: Texte La Conférence Donnée En Décembre, pp. 1–24 (2010)
30. Benndorf, V., Normann, H.T.: The willingness to sell personal data. Scand. J. Econ. **120**, 1260–1278 (2018)
31. Li, C., Li, D.Y., Miklau, G., Suciu, D.: A theory of pricing private data. ACM Trans. Database Syst. (TODS) **39**(4), 1–28 (2014). Article no. 34
32. Preibusch, S.: The value of privacy in web search. In: WEIS (2013)
33. Cofone, I.N.: The value of privacy: keeping the money where the mouth is. In: 14th Annual Workshop on the Economics of Information Security, pp. 1–31 (2015)
34. Malgieri, G., Custers, B.: Pricing privacy - the right to know the value of your personal data. Comput. Law Secur. Rev. **34**(2), 289–303 (2018)
35. Cranor, L.F., Arjula, M., Guduru, P.: Use of a P3P user agent by early adopters. In: Proceedings of the ACM Workshop on Privacy in the Electronic Society, WPES 2002, pp. 1–10 (2002)
36. Federrath, H.: Privacy enhanced technologies: methods – markets – misuse. In: Katsikas, S., López, J., Pernul, G. (eds.) TrustBus 2005. LNCS, vol. 3592, pp. 1–9. Springer, Heidelberg (2005). https://doi.org/10.1007/11537878_1
37. Schomakers, E.M., Lidynia, C., Vervier, L., Ziefle, M.: Of guardians, cynics, and pragmatists - a typology of privacy concerns and behavior. In: IoTBDS, pp. 153–163 (2018)
38. Roßnagel, H., Zibuschka, J., Pimenides, L., Deselaers, T.: Facilitating the adoption of tor by focusing on a promising target group. In: Jøsang, A., Maseng, T., Knapskog, S.J. (eds.)

NordSec 2009. LNCS, vol. 5838, pp. 15–27. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-04766-4_2
39. Böhme, R., Koble, S.: On the viability of privacy-enhancing technologies in a self-regulated business-to-consumer market: will privacy remain a luxury good? Dresden (2007)
40. Wilcoxon, F.: Individual comparisons by ranking methods. Biom. Bull. **1**, 80–83 (1945)
41. Mann, H.B., Whitney, D.R.: On a test of whether one of two random variables is stochastically larger than the other. Ann. Math. Stat. **18**, 50 (1947)
42. Benjamini, Y.: Opening the box of a boxplot. Am. Stat. **42**, 257–262 (1988)
43. McKelvey, D., Zavorina, W.: A statistical model for the analysis of ordinal level dependent variables. J. Math. Sociol. **4**, 103–120 (1975)
44. Donthu, N., Gilliland, D.: Observations: the infomercial shopper. J. Advert. Res. **36**, 69–76 (1996)
45. Frik, A., Gaudeul, A.: The relation between privacy protection and risk attitudes, with a new experimental method to elicit the implicit monetary value of privacy. CEGE discussion papers, Number 296. SSRN (2016). http://papers.ssrn.com/abstract=2874202
46. Christofides, E., Muise, A., Desmarais, S.: Risky disclosures on Facebook: the effect of having a bad experience on online behavior. J. Adolesc. Res. **27**, 714–731 (2012)
47. Pavlou, P.A.: Consumer acceptance of electronic commerce: integrating trust and risk with the technology acceptance model. Int. J. Electron. Commer. **7**, 101–134 (2003)
48. Harborth, D., Pape, S.: Explaining technology use behaviors of privacy-enhancing technologies: the case of Tor and JonDonym. Submitted to IEEE European Symposium on Security and Privacy (EuroS&P 2019) (2019)
49. Harborth, D., Pape, S.: German translation of the concerns for information privacy (CFIP) construct. SSRN (2018). https://ssrn.com/abstract=3112207
50. Schmitz, C.: LimeSurvey Project Team. http://www.limesurvey.org. Accessed 12 Dec 2018

C.8 How Privacy Concerns and Trust and Risk Beliefs Influence Users' Intentions to Use Privacy-Enhancing Technologies – The Case of Tor

David Harborth and Sebastian Pape. How privacy concerns and trust and risk beliefs influence users' intentions to use privacy-enhancing technologies – the case of Tor. In *52nd Hawaii International Conference on System Sciences (HICSS) 2019*, pages 4851–4860, 01 2019. doi: 10125/59923. URL https://scholarspace.manoa.hawaii.edu/bitstream/10125/59923/1/0483.pdf

Proceedings of the 52nd Hawaii International Conference on System Sciences | 2019

How Privacy Concerns and Trust and Risk Beliefs Influence Users' Intentions to Use Privacy-Enhancing Technologies - The Case of Tor

David Harborth
Chair of Mobile Business and Multilateral Security
Goethe University Frankfurt am Main
david.harborth@m-chair.de

Sebastian Pape
Chair of Mobile Business and Multilateral Security
Goethe University Frankfurt am Main
sebastian.pape@m-chair.de

Abstract

Due to an increasing collection of personal data by internet companies and several data breaches, research related to privacy gained importance in the last years in the information systems domain. Privacy concerns can strongly influence users' decision to use a service. The Internet Users Information Privacy Concerns (IUIPC) construct is one operationalization to measure the impact of privacy concerns on the use of technologies. However, when applied to a privacy enhancing technology (PET) such as an anonymization service the original rationales do not hold anymore. In particular, an inverted impact of trusting and risk beliefs on behavioral intentions can be expected. We show that the IUIPC model needs to be adapted for the case of PETs. In addition, we extend the original causal model by including trust beliefs in the anonymization service itself. A survey among 124 users of the anonymization service Tor shows that they have a significant effect on the actual use behavior of the PET.

1. Introduction

"Surveillance is the business model of the internet. Everyone is under constant surveillance by many companies, ranging from social networks like Facebook to cellphone providers." [1]. Privacy and the related concerns have been discussed since the very beginning of computer sharing [2]. Due to a raising economic interest in personal data during the last years [3], privacy gains an increasing importance in individuals' everyday life. The majority of internet users has privacy concerns and feels a strong need to protect their privacy [4].

A popular model for measuring and explaining privacy concerns of online users is the model focusing on the Internet Users Information Privacy Concerns (IUIPC) construct by Malhotra et al. [5]. Their research involves a theoretical framework and an instrument for operationalizing privacy concerns, as well as a causal model for this construct including trust and risk beliefs about the online companies' data handling of personal information. The IUIPC construct has been used in various contexts, e.g. Internet of Things [6], internet transactions [7] and Mobile Apps [8]. Originally, the IUIPC instrument was applied to use cases for individuals' decisions to disclose personal information to service providers. However, for privacy enhancing technologies (PETs) the primary purpose is to help users to protect personal information when using regular internet services. As a consequence, it is necessary to reconsider the impact of trust and risk beliefs within IUIPC's causal model with respect to PETs. We expected this impact to be inverted and thus the trust model needs to be adapted for the investigation of PETs. In addition, trust in the PET itself is an important factor to consider. This is the case since Tor is used by a diverse group of people whose life might be endangered in case their identity is revealed (e.g. whistleblowers, opposition supporters, etc. [9]). To the best of our knowledge the IUIPC construct has never been applied to a PET. Thus, we address the following research questions:

1. *What influence have privacy concerns and associated trust and risk beliefs on the behavioral intention and actual use of Tor?*
2. *What influence does trust in Tor itself have on the behavioral intention and the actual use?*

For that purpose, we conducted an online survey with users of one of the most widely used anonymization services Tor (Tor has approximately 2,000,000 regular users) [9]. We collected 124 complete questionnaires out of 314 participants for the empirical analysis. Our results contribute to the understanding of users' perceptions about PETs and indicate how privacy concerns and trust and risk beliefs influence the use behavior of PETs.

The remainder of the paper is as follows: Sect. 2 introduces Tor and lists related work on PETs. In Sect. 3, we present research hypotheses and the data collection process. We assess the reliability and validity of our results in Sect. 4. In Sect. 5, we discuss the implications and limitations of our work and suggest future work.

This research was partly funded by the German Federal Ministry of Education and Research (BMBF) with grant number: 16KIS0371.

URI: https://hdl.handle.net/10125/59923
ISBN: 978-0-9981331-2-6

2. Background and Related Work

Privacy-Enhancing Technologies (PETs) is an umbrella term for different privacy protecting technologies. PETs can be defined as a "coherent system of ICT measures that protects privacy [...] by eliminating or reducing personal data or by preventing unnecessary and/or undesired processing of personal data; all without losing the functionality of the data system" [10, p. 1].

In this paper, we investigate the privacy, trust and risk beliefs associated with PETs for the case of the anonymity service Tor [9]. Tor is a free-to-use anonymity service that is based on the onion routing principle. Everybody can operate a server (relay) over which the encrypted traffic is routed. The routing occurs randomly over several different servers distributed world-wide. Tor aims to protect against an adversary who can observe or control some fraction of network traffic, but it does not protect against a global passive adversary, which means an adversary who can observe all network connections. Among the available PETs, Tor has one of the biggest user bases with approximately 2,000,000 active users [9].

Related work on PETs considers mainly usability studies and does not primarily focus on privacy concerns and related trust and risk beliefs of PET users. For example, Lee et al. [11] assess the usability of the Tor Launcher and propose recommendations to overcome the found usability issues. Benenson et al. [12] investigate acceptance factors for anonymous credentials. Among other things, they find that trust in the PET has no statistically significant impact on the intention to use the service. This result is relevant for our study since we hypothesize that trust in Tor has a positive effect on the actual use of the service (see Section 3.1). Another highly relevant study for our research is the one by Brecht et al. [13], who investigate acceptance factors of anonymization services. Among other variables, they hypothesize a positive influence of privacy concerns on the intention to use such a service. Although they find a statistically significant effect, the effect is relatively small (effect size of 0.061) compared to other variables like perceived usefulness or internet privacy awareness. In contrast to our study, Brecht et al. [13] use another operationalization of privacy concerns (the one by Dinev and Hart [14]) and they do not investigate it in the nomological network with trust and risk beliefs.

3. Methodology

We base our research on the Internet Users Information Privacy Concerns (IUIPC) model by Malhotra et al. [5]. The original research on this model investigates the role of users' information privacy concerns in the context of releasing personal information to a marketing service provider. Since we are focusing on the role of privacy concerns, trust and risk beliefs for the case of a PET (i.e. Tor), we adapt the original model according to the following logic. Originally, the service in question can be seen as the attacker (from a privacy point of view). If we apply the model to a service with the opposite goal, namely protecting the privacy of its users, certain relationships need to change. We will elaborate on the detailed changes in the next section. In addition, to this we extend the original model by trusting beliefs in the PET itself. We argue that the level of trust in a PET is a crucial factor determining the use decision.

For analyzing the cause-effect relationships between the latent (unobserved) variables, we use structural equation modelling (SEM). Since our research goal is to predict the target constructs behavioral intention and actual use behavior of Tor, we use partial least squares SEM (PLS-SEM) for our analysis [15, 16] and not covariance-based SEM. In the following subsections, we discuss the hypotheses based on the IUIPC model [5], the questionnaire and the data collection process.

3.1. Research Hypotheses

The structural model contains several relationships between exogenous and endogenous variables (cf. Fig. 1). We develop our research hypotheses for these relationships along the hypotheses of the IUIPC model [5]. IUIPC is operationalized as a second-order construct[1] of the sub-constructs collection (COLL), awareness (AWA) and control (CONTROL). Thus, the users' privacy concerns are determined by their concerns about "[...] individual-specific data possessed by others relative to the value of benefits receive" [5, p. 338], the control they have over their own data (i.e. possibilities to change or opt-out) and the "[...] degree to which a consumer is concerned about his/her awareness of organizational information privacy practices" [5, p. 339].

The effect of IUIPC on the behavioral intention is mediated by trusting beliefs and risk beliefs. Trusting beliefs are users' perceptions about the behavior of online firms to protect the users' personal information. In contrast, risk beliefs represent users' perception about losses associated with providing personal data to online firms [5]. Thus, the higher the privacy concerns of a user, the lower are his or her trusting beliefs and the higher are his or her risk beliefs. In addition, a higher level of trust is assumed to decrease the risk beliefs. Thus, we hypothesize:

[1] Due to space limitations, we will not elaborate on second-order constructs in more detail. For an extensive discussion see Steward [17].

1. *Internet Users Information Privacy Concerns (IUIPC) have a negative effect on Trusting Beliefs (TB).*
2. *Internet Users Information Privacy Concerns (IUIPC) have a positive effect on Risk Beliefs (RB).*
3. *Trusting Beliefs (TB) have a negative effect on Risk Beliefs (RB).*

Since we investigate the use of a specific PET, we extend the model by the trust in Tor itself with the adapted trust construct by Pavlou [18]. However, in order to protect their privacy, users with higher privacy concerns are assumed to rather trust the privacy-enhancing technology compared to online firms which process personal data. This is especially true, because we surveyed users of a PET which are assumed to take great care of their privacy. Therefore, we hypothesize:

4. *Internet Users Information Privacy Concerns (IUIPC) have a positive effect on the trusting beliefs in Tor (TB_{Tor}).*

Trust is an important factor in the acceptance decision of users [18]. Mcknight et al. [19] show that trust in a specific technology will positively affect individuals intention to explore the technology and to use more features of the technology in a postadoption context. Especially for the case of privacy protection, we assume that trust in the technology is a major factor for the intention to use the technology. For a further discussion on the concept of trust in a technology, we refer to Lankton et al. [20]. We hypothesize that:

5. *Trusting beliefs in Tor (TB_{Tor}) have a positive effect on the behavioral intention to use Tor (BI).*

It is logical that trusting beliefs have a positive effect and risk beliefs have a negative effect on releasing data and thus the intended behavior of using a regular service. However, for use behavior of a PET, we assume these effects reverse. The higher the trusting beliefs in online firms, the lower is the use frequency of Tor, since the protection of data becomes less important. Following this rationale, a higher degree of risk beliefs in data processing of online firms leads to a higher degree of use. Thus, we hypothesize that:

6. *Trusting beliefs (TB) have a negative effect on the behavioral intention to use Tor (BI).*
7. *Risk beliefs (RB) have a positive effect on the behavioral intention to use Tor (BI).*

Research on the relationship between behavioral intention and use behavior goes back to Fishbein et al. [21]. Later research indicates a positive link between the two constructs [22]. Thus, we hypothesize that:

8. *The behavioral intention to use Tor (BI) has a positive effect on the actual use behavior (USE).*

3.2. Data Collection

The questionnaire constructs are adapted from the original IUIPC paper [5]. We conducted the study with German and English speaking Tor users. Thus, we administered two questionnaires. All items for the German questionnaire had to be translated into German since all of the constructs are adapted from English literature. To ensure content validity of the translation, we followed a rigorous translation process. First, we translated the English questionnaire into German with the help of a certified translator (translators are standardized following the DIN EN 15038 norm). The German version was then given to a second independent certified translator who retranslated the questionnaire to English. This step was done to ensure the equivalence of the translation. Third, a group of five academic colleagues checked the two English versions with regard to this equivalence. All items were found to be equivalent. The items of the English version can be found in Appendix B.

Since we investigate the effect of privacy concerns, trust and risk beliefs on the use of Tor, we collected data of actual users. We installed the surveys on a university server and managed it with the survey software LimeSurvey (version 2.72.6) [23]. The links to the English and German version were distributed over multiple channels on the internet. Although there are approximately 2,000,000 active users of the service, it was relatively difficult to gather the necessary number of complete answers for a valid and reliable quantitative analysis. Thus, to foster future research about Tor users, we provide an overview of every distribution channel in the Appendix A. In sum, 314 participants started the questionnaire (245 for the English version, 40 for the English version posted in hidden service forums and 29 for the German version). Of those 314 approached participants, 135 (105 for the English version, 13 for the English version posted in hidden service forums and 17 for the German version) filled out the questionnaires completely. After deleting all sets from participants who answered a test question in the middle of the survey incorrectly, 124 usable data sets remained for the following analysis.

The demographic questions were not mandatory to fill out. This was done on purpose since we assumed that most of the participants are highly sensitive with respect to their personal data. Therefore, we had to resign from a discussion of the demographics in our research context. This decision is backed up by Singh and Hill, who found no statistically significant differences across gender, income groups, educational levels, or political affiliation in the desire to protect one's privacy [4].

4. Results

We tested the model using SmartPLS version 3.2.7 [24]. Before looking at the result of the structural model and discussing its implications, we discuss the measurement model, and check for the reliability and validity of our results. This is a precondition of being able to interpret the results of the structural model. Furthermore, it is recommended to report the computational settings. For the PLS algorithm, we choose the path weighting scheme with a maximum of 300 iterations and a stop criterion of 10^{-7}. For the bootstrapping procedure, we use 5000 bootstrap subsamples and no sign changes as the method for handling sign changes during the iterations of the bootstrapping procedure.

4.1. Assessment of the Measurement Model

As the model is measured solely reflectively, we need to evaluate the internal consistency reliability, convergent validity and discriminant validity to assess the measurement model properly [15].

Internal Consistency Reliability Internal consistency reliability (ICR) measurements indicate how well certain indicators of a construct measure the same latent phenomenon. Two standard approaches for assessing ICR are Cronbach's α and the composite reliability. The values of both measures should be between 0.7 and 0.95 for research that builds upon accepted models. Values of Cronbach's α are seen as a lower bound and values of the composite reliability as an upper bound of the assessment [16]. Table 1 includes the ICR of the variables in the last two rows. It can be seen that all values for Cronbach's α are above the lower threshold of 0.7 except for RISK. However, for the composite reliability the value for RISK is higher than 0.7. Therefore, we argue that ICR is not a major issue for this variable. For all variables, no value is above 0.95. Values above that upper threshold indicate that the indicators measure the same dimension of the latent variable, which is not optimal with regard to the validity [16]. In sum, ICR is established for our variables. Since IUIPC and USE are single-item constructs they have ICR values of 1.

Convergent Validity Convergent validity determines the degree to which indicators of a certain reflective construct are explained by that construct. This is assessed by calculating the outer loadings of the indicators of the constructs (indicator reliability) and by looking at the average variance extracted (AVE) [15]. Loadings above 0.7 imply that the indicators have much in common, which is desirable for reflective measurement models [16]. Table 1 shows the outer loadings in bold on the diagonal. All loadings were higher than 0.7, except for TRUST4 with a value of 0.275. Therefore, we dropped this item after an initial analysis. Convergent validity for the construct is assessed by the AVE. AVE is equal to the sum of the squared loadings divided by the number of indicators. A threshold of 0.5 is acceptable, indicating that the construct explains at least half of the variance of the indicators [16]. The diagonal values of Table 2 present the AVE of our constructs. All values are well above 0.5, demonstrating convergent validity.

Discriminant Validity Discriminant validity measures the degree of uniqueness of a construct compared to other constructs. Comparable to the convergent validity assessment, two approaches are used for investigating discriminant validity. The first approach, assessing cross-loadings, is dealing with single indicators. All outer loadings of a certain construct should be larger than its cross-loadings with other constructs [15]. Table 1 illustrates the cross-loadings as off-diagonal elements. All cross-loadings are smaller than the outer loadings, fulfilling the first assessment approach of discriminant validity. The second approach is on the construct level and compares the square root of the constructs' AVE with the correlations with other constructs. The square root of the AVE of a single construct should be larger than the correlation with other constructs (Fornell-Larcker criterion) [16]. Table 2 contains the square root of the AVE on the diagonal in parentheses. All values are larger than the correlations with other constructs, indicating discriminant validity. Since there are problems in determining the discriminant validity with both approaches, researchers propose the heterotrait-monotrait ratio (HTMT) for assessing discriminant validity as a superior approach [25]. HTMT divides between-trait correlations by within-trait correlations, therefore providing a measure of what the true correlation of two constructs would be if the measurement is flawless [16]. Values close to 1 for HTMT indicate a lack of discriminant validity. A conservative threshold is 0.85 [25]. Table 3 contains the values for HTMT and no value, except for the correlation between IUIPC and COLL (with 0.888), is above the suggested threshold of 0.85. To assess if the HTMT statistics are significantly different from 1, we conducted a bootstrapping procedure with 5,000 subsamples to get the confidence interval in which the true HTMT value lies with a 95% chance. The HTMT measure requires that no confidence interval contains the value 1. The conducted analysis shows that this is the case, and thus discriminant validity is established for our model.

Common Method Bias The common method bias (CMB) can occur if data is gathered with a self-reported survey at one point in time in one questionnaire [26]. Since this is the case in our research design, the need to

Table 1. Loadings and Cross-Loadings of the Reflective Items and Internal Consistency Reliability

Construct	AWA	CONTROL	COLL	RB	TB	TB_{Tor}	BI	IUIPC	USE
AWA1	**0.911**	0.234	0.302	0.223	-0.136	0.066	0.202	0,630	-0.124
AWA2	**0.923**	0.230	0.219	0.136	-0.155	0.072	0.198	0.586	-0.171
AWA3	**0.891**	0.323	0.315	0.221	-0.103	0.066	0.250	0.660	-0.059
CONTROL1	0.095	**0.825**	0.271	0.106	-0.167	0.137	0.215	0.475	-0.021
CONTROL2	0.405	**0.821**	0.226	0.245	-0.156	0.132	0.237	0.577	-0.033
CONTROL3	0.174	**0.756**	0.438	0.214	-0.345	0.098	0.099	0.578	0.068
COLL1	0.264	0.358	**0.888**	0.547	-0.468	0.176	0.301	0.742	0.045
COLL2	0.206	0.332	**0.812**	0.205	-0.335	0.232	0.376	0.665	0.042
COLL3	0.292	0.359	**0.906**	0.444	-0.446	0.272	0.376	0.764	0.071
COLL4	0.304	0.309	**0.850**	0.467	-0.403	0.182	0.316	0.720	0.091
RB1	0.196	0.200	0.487	**0.880**	-0.453	0.217	0.258	0.429	-0.015
RB2	0.170	0.160	0.326	**0.831**	-0.298	0.156	0.233	0.312	0.015
RB3	0.155	0.252	0.364	**0.857**	-0.354	0.233	0.221	0.359	0.007
RB4	0.245	0.231	0.374	**0.827**	-0.260	0.257	0.326	0.396	0.042
RB5	-0.105	-0.145	-0.427	**-0.702**	0.401	-0.004	-0.144	-0.339	0.003
TB1	-0.149	-0.261	-0.455	-0.417	**0.898**	-0.097	-0.265	-0.412	-0.050
TB2	-0.118	-0.186	-0.410	-0.377	**0.887**	-0.033	-0.194	-0.347	-0.109
TB3	-0.107	-0.339	-0.397	-0.395	**0.775**	-0.131	-0.155	-0.387	-0.007
TB5	-0.069	-0.009	-0.219	-0.070	**0.663**	-0.109	-0.169	-0.158	-0.007
$TB_{Tor}1$	0.064	0.149	0.257	0.159	-0.087	**0.879**	0.561	0.225	-0.050
$TB_{Tor}2$	0.077	0.121	0.236	0.244	-0.124	**0.925**	0.554	0.209	-0.020
$TB_{Tor}3$	0.059	0.138	0.169	0.178	-0.079	**0.883**	0.488	0.169	0.002
BI1	0.236	0.240	0.355	0.228	-0.249	0.586	**0.865**	0.384	0.166
BI2	0.262	0.202	0.322	0.319	-0.152	0.465	**0.859**	0.363	0.075
BI3	0.143	0.158	0.363	0.234	-0.233	0.522	**0.923**	0.323	0.216
IUIPC	0.691	0.685	0.837	0.451	-0.431	0.226	0.404	**1,000**	-0.009
USE	-0.128	0.008	0.073	0.010	-0.059	-0.026	0.177	-0.009	**1,000**
Cronbach's α	0.894	0.722	0.887	0.567	0.831	0.877	0.859	1.000	1.000
Comp. Reliability	0.934	0.843	0.922	0.817	0.884	0.924	0.914	1.000	1.000

test for CMB arises. An unrotated principal component factor analysis is performed with the software package STATA 14.0 to conduct the Harman's single-factor test to address the issue of CMB [27]. The assumptions of the test are that CMB is not an issue if there is no single factor that results from the factor analysis or that the first factor does not account for the majority of the total variance [27]. The test shows that seven factors have eigenvalues larger than 1 which account for 75.35% of the total variance. The first factor explains 30.29% of the total variance. Based on the results of previous literature [28], we argue that CMB is not likely to be an issue in the data set.

4.2. Assessment and Results of the Structural Model

To assess the structural model, we follow the steps proposed by Hair et al. [16] which include an assessment of possible collinearity problems, of path coefficients, of the level of R^2, of the effect size f^2, of the predictive relevance Q^2 and the effect size q^2. We address these evaluation steps to ensure the predictive power of the model with regard to the target constructs.

Collinearity Collinearity is present if two predictor variables are highly correlated with each other. To address this issue, we assess the inner variance inflation factor (VIF). All VIFs above 5 indicate that collinearity between constructs is present. For our model, the highest VIF is 1.278. Thus collinearity is apparently not an issue.

Significance and Relevance of Model Relationships Figure 1 shows the results of the path estimations and the R^2-values of the endogenous variables BI and USE. The R^2 is 0.400 for BI and 0.031 for USE. Thus, our models explains 40% of the variance of BI and 3.1% of USE. There are different proposals for interpreting the size of this value. We choose to use the very conservative threshold proposed by Hair et al. [15], where R^2 values are weak with values around 0.25, moderate with 0.50 and substantial with 0.75. Based on this classification, the R^2 value for BI is weak to moderate and for USE the value is very weak. For use behavior

Table 2. Discriminant Validity with AVEs and Construct Correlations

Constructs (AVE)	AWA	BI	COLL	CONTROL	IUIPC	RB	TB	TB_{Tor}	USE
AWA (0.825)	0.908								
BI (0.780)	0.240	0.883							
COLL (0.748)	0.309	0.395	0.865						
CONTROL (0.642)	0.291	0.228	0.393	0.801					
IUIPC (1.000)	0.691	0.404	0.837	0.685	1,000				
RB (0.675)	0.215	0.291	0.486	0.242	0.451	0.822			
TB (0.658)	-0.143	-0.244	-0.480	-0.283	-0.431	-0.434	0.811		
TB_{Tor} (0.803)	0.075	0.599	0.249	0.152	0.226	0.216	-0.109	0.896	
USE (1.000)	-0.128	0.177	0.073	0.008	-0.009	0.010	-0.059	-0.026	1,000

Note: AVEs in parentheses in the first column. Values for $\sqrt{AVE}$ are shown on the diagonal and construct correlations are off-diagonal elements.

Table 3. Heterotrait-Monotrait Ratio (HTMT)

Constructs	AWA	BI	COLL	CONTROL	IUIPC	RB	TB	TB_{Tor}	USE
BI	0.274								
COLL	0.343	0.452							
CONTROL	0.346	0.290	0.486						
IUIPC	0.728	0.436	0.888	0.798					
RB	0.238	0.337	0.541	0.294	0.478				
TB	0.159	0.278	0.528	0.336	0.439	0.449			
TB_{Tor}	0.084	0.681	0.280	0.192	0.240	0.244	0.131		
USE	0.138	0.186	0.077	0.060	0.009	0.021	0.058	0.029	

several participants answered that they never use Tor (21 participants answered "never") although they stated to use the service several years (answers to the question: How many years are you using Tor? with a median of 6 years and an average of 6.87 years on a seven-point Likert scale). The correlation coefficient between the years of using Tor and the use frequency is very small, negative and statistically insignificant with -0.0222 and a p-value of 0.8066. These 21 answers massively bias the results for the relationship between behavioral intention and actual use behavior (the median value of use frequency is 5). However, we cannot explain why the participants answered like this. They either misunderstood the question, answered it intentionally like this to disguise their activity with Tor or found the scale for use behavior inappropriate. This might be due to the fact that the scale only contains "once a month" as the lowest use frequency besides "never". It might be possible that these 21 users use Tor only a few times per year or that they used Tor some years ago and have not used it again since then. Therefore, they might have chosen never as an answer. However, we used an established scale to measure use behavior [29], but recommend to consider this issue in future research studies with a similar context.

The path coefficients are presented on the arrows connecting the exogenous and endogenous constructs in Figure 1. Statistical significance is indicated by asterisks, ranging from three asterisks for p-values smaller than 0.01 to one asterisk for p-values smaller than 0.10. We chose this p-value range since p-values tend to be larger if the sample size is comparable small and we wanted to capture also significant effects above the 5% level. The p-value indicates the probability that a path estimate is incorrectly assumed to be significant. Thus, the lower the p-value, the higher the probability that the given relationship exists. The relevance of the path coefficients is shown by the relative size of the coefficient compared to the other explanatory variables [16].

It can be seen that IUIPC has a relatively large statistically significant negative effect on trusting beliefs and a positive effect on risk beliefs. The effect of IUIPC on trusting beliefs in Tor is significant, positive and relatively weak compared to the other significant effects in the model. The construct trusting beliefs has a statistically significant medium-sized negative effect on risk beliefs. The effects of trusting beliefs and risk beliefs on behavioral intention are not statistically significant (for both $p \geq 0.10$). In contrast, the effect of trusting beliefs in Tor on behavioral intention is highly statistically significant, positive and large with 0.560.

Effect Sizes f^2 The f^2 effect size measures the impact of a construct on the endogenous variable by omitting it from the analysis and assessing the resulting change in the R^2 value [16]. The values are assessed based on thresholds by Cohen [30], who defines effects as small, medium and large for values of 0.02, 0.15 and 0.35, respectively. Table 4 shows the results of the f^2 evaluation. Values in italics indicate small effects, values

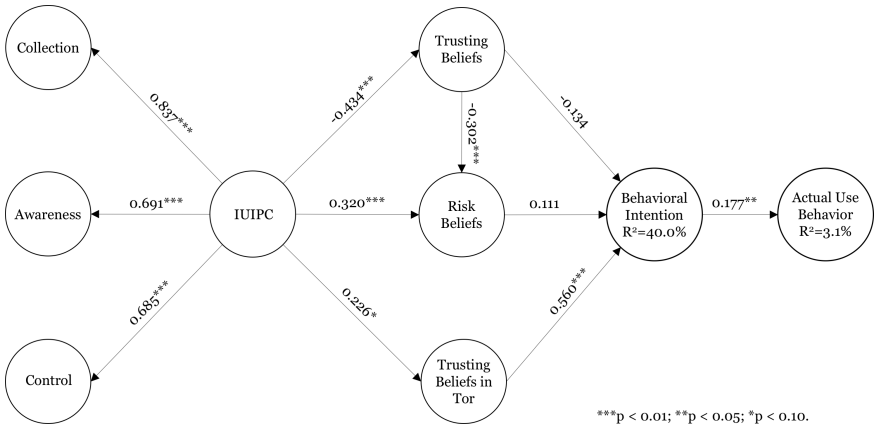


Figure 1. Path Estimates and Adjusted R^2 values of the Structural Model

Table 4. f^2 and q^2 Effect Size Assessment Values

Variables	f^2	q^2
Endogenous / Exogenous	BI	BI
RB	0.016	0.018
TB	*0.025*	*0.025*
TB_{Tor}	***0.499***	***0.766***

in bold indicate medium effects and values in bold and italics indicate large effects. All other values have no substantial effect. The results correspond to those of the previous analysis of the path coefficients whereas trusting beliefs have a small effect on the behavioral intention to use tor. As the path estimates have shown, trust in tor has a large effect on the behavioral intention.

Predictive Relevance Q^2 The Q^2 measure indicates the out-of-sample predictive relevance of the structural model with regard to the endogenous latent variables based on a blindfolding procedure [16]. We used an omission distance d=7. Recommended values for d are between five and ten [15]. Furthermore, we report the Q^2 values of the cross-validated redundancy approach, since this approach is based on both the results of the measurement model as well as of the structural model [16]. Detailed information about the calculation cannot be provided due to space limitations. For further information see Chin [31]. Values above 0 indicate that the model has the property of predictive relevance. In our case, the Q^2 value is equal to 0.278 for BI and 0.002 for USE. Since they are larger than zero, predictive relevance of the model is established.

Effect Sizes q^2 The assessment of q^2 follows the same logic as the one of f^2. It is based on the Q^2 values of the endogenous variables and calculates the individual predictive power of the exogenous variables by omitting them and comparing the change in Q^2 [16]. All individual values for q^2 are calculated with an omission distance d of seven. The results are shown in Table 4. The thresholds for the f^2 interpretation can be applied here, too [30]. Values in italics indicate small effects and values in bold indicate medium effects. All other values have no substantial effect. As before, only the trust in Tor has a large effect, implying the highest predictive power of all included exogenous variables.

5. Discussion and Conclusion

Based on our results, hypotheses H1 to H5 and H8 can be confirmed, whereas H6 and H7 cannot be confirmed (cf. Table 5). The results for H6 and H7 are surprising, considering that they are in contrast to the rationale explained in Sect. 3.1 and the results from previous literature [5]. However, it must be said that when effect sizes are rather small it is possible that the relatively small sample size of 124 leads to a statistical non-significance. Thus, we cannot rule out that the effects of risk beliefs and trusting beliefs on behavioral intention would be significant with a larger sample size. Thus, only the degree of trust in the PET (Tor) has a direct significant effect on the intention to use the PET. This result shows that a reputation of being trustworthy is crucial for a PET provider. The trusting beliefs in the PET itself are positively influenced by the users' information privacy concerns. Thus, the results imply that users with a higher level of privacy concerns rather tend to trust a PET.

The limitations of the study primarily concern the sample composition and size. First, a larger sample

Table 5. Summary of the Results

	Hypothesis	Result
H1:	Internet Users Information Privacy Concerns (IUIPC) have a negative effect on Trusting Beliefs (TB)	✓
H2:	Internet Users Information Privacy Concerns (IUIPC) have a positive effect on Risk Beliefs (RB)	✓
H3:	Trusting Beliefs (TB) have a negative effect on Risk Beliefs (RB)	✓
H4:	Internet Users Information Privacy Concerns (IUIPC) have a positive effect on the trusting beliefs in Tor (TB_{Tor})	✓
H5:	Trusting beliefs in Tor (TB_{Tor}) have a positive effect on the behavioral intention to use Tor (BI)	✓
H6:	Trusting beliefs (TB) have a negative effect on the behavioral intention to use Tor (BI)	✗
H7:	Risk beliefs (RB) have a positive effect on the behavioral intention to use Tor (BI)	✗
H8:	The behavioral intention to use Tor (BI) has a positive effect on the actual use behavior (USE)	✓

would have been beneficial. However, in general, a sample of 124 participants is acceptable for our kind of statistical analysis [16] and active users of a PET are hard to find for a relatively long online questionnaire. This is especially the case, if they do not have any financial rewards as in our study and if they are highly privacy sensitive which might repel them to disclose any kind of information (even if it is anonymous). Second, the combination of the results of the German and the English questionnaire can be a potential source of errors. German participants might have understood questions differently than the English participants. We argue that we achieved equivalence with regard to the meaning through conducting a thorough translation process, and therefore limiting this potential source of error to the largest extent possible. In addition, combining the data was necessary from a pragmatic point of view to get a sample size as large as possible for the statistical analysis. Lastly, possible self-report biases (e.g. social desirability) might exist. We addressed this possible issue by gathering the data fully anonymized. As discussed earlier, we had issues with certain data sets of participants with regard to actual use behavior (cf. Sect. 4.2.). Although it might be more beneficial in certain settings to directly refer to actual use behavior as the sole target variable, we decided to include behavioral intention as an antecedent because of these issues.

Further work is required to investigate the specific determinants of use decisions for or against PETs and break down the interrelationships between the associated antecedents. In particular, it would be interesting to investigate the relationship between trusting beliefs in online companies and trust in the PET itself. A theoretical underlying is required to include this relationship in our structural equation model.

In this paper, we contributed to the research on privacy-enhancing technologies and users' privacy by assessing the specific relationships between information privacy concerns, trusting beliefs in online firms and a privacy-enhancing technology (in our case Tor), risk beliefs associated with online firms data processing and the actual use behavior of Tor. By adapting and extending the IUIPC model by Malhotra et al. [5], we could show that several of the assumptions for regular online services do not hold for PETs.

References

[1] L. Mineo, "On internet privacy, be very afraid (Interview with Bruce Schneier)." https://news.harvard.edu/gazette/story/2017/08/when-it-comes-to-internet-privacy\-be-very-afraid-analyst-suggests/, 08 2017.

[2] E. E. David and R. M. Fano, "Some thoughts about the social implications of accessible computing," in *Proceedings 1965 Fall Joint Computer Conference*, 1965. Available via http://www.multicians.org/fjcc6.html.

[3] M. Bédard, "The underestimated economic benefits of the internet," regulation series, The Montreal Economic Institute, 2016. Economic Notes.

[4] T. Singh and M. E. Hill, "Consumer privacy and the Internet in Europe: a view from Germany," *Journal of consumer marketing*, vol. 20, no. 7, pp. 634–651, 2003.

[5] N. K. Malhotra, S. S. Kim, and J. Agarwal, "Internet users' information privacy concerns (IUIPC): The construct, the scale, and a causal model," *Information Systems Research*, vol. 15, pp. 336–355, dec 2004.

[6] P. E. Naeini, S. Bhagavatula, H. Habib, M. Degeling, L. Bauer, L. Cranor, and N. Sadeh, "Privacy expectations and preferences in an iot world," in *Symposium on Usable Privacy and Security (SOUPS)*, 2017.

[7] J. Heales, S. Cockcroft, and V.-H. Trieu, "The influence of privacy, trust, and national culture on internet transactions," in *Social Computing and*

Social Media. Human Behavior (G. Meiselwitz, ed.), (Cham), pp. 159–176, Springer International Publishing, 2017.

[8] F. Raber and A. Krueger, "Towards understanding the influence of personality on mobile app permission settings," in *IFIP Conference on Human-Computer Interaction*, pp. 62–82, 2017.

[9] The Tor Project. https://www.torproject.org, 2018.

[10] J. J. Borking and C. Raab, "Laws, PETs and Other Technologies for Privacy Protection," *Journal of Information, Law and Technology*, vol. 1, pp. 1–14, 2001.

[11] L. Lee, D. Fifield, N. Malkin, G. Iyer, S. Egelman, and D. Wagner, "A Usability Evaluation of Tor Launcher," *Proceedings on Privacy Enhancing Technologies*, no. 3, pp. 90–109, 2017.

[12] Z. Benenson, A. Girard, and I. Krontiris, "User Acceptance Factors for Anonymous Credentials: An Empirical Investigation," *14th Annual Workshop on the Economics of Information Security (WEIS)*, pp. 1–33, 2015.

[13] F. Brecht, B. Fabian, S. Kunz, and S. Mueller, "Are You Willing to Wait Longer for Internet Privacy?," in *ECIS 2011 Proceedings*, 2011.

[14] T. Dinev and P. Hart, "An extended privacy calculus model for e-commerce transactions," *Information Systems Research*, vol. 17, no. 1, pp. 61–80, 2006.

[15] J. Hair, C. M. Ringle, and M. Sarstedt, "PLS-SEM: Indeed a Silver Bullet," *The Journal of Marketing Theory and Practice*, vol. 19, no. 2, pp. 139–152, 2011.

[16] J. Hair, G. T. M. Hult, C. M. Ringle, and M. Sarstedt, *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)*. SAGE Publications, 2017.

[17] K. A. Stewart and A. H. Segars, "An Empirical Examination of the Concern for Information Privacy Instrument," *Information Systems Research*, vol. 13, no. 1, pp. 36–49, 2002.

[18] P. A. Pavlou, "Consumer Acceptance of Electronic Commerce: Integrating Trust and Risk with the Technology Acceptance Model," *International Journal of Electronic Commerce*, vol. 7, no. 3, pp. 101–134, 2003.

[19] D. H. Mcknight, M. Carter, J. B. Thatcher, and P. F. Clay, "Trust in a specific technology: An investigation of its components and measures," *ACM Transactions on Management Information Systems (TMIS)*, vol. 2, no. 2, p. 12, 2011.

[20] N. K. Lankton, D. H. McKnight, and J. Tripp, "Technology, humanness, and trust: Rethinking trust in technology," *Journal of the Association for Information Systems*, vol. 16, no. 10, p. 880, 2015.

[21] M. Fishbein and I. Ajzen, *Belief, Attitude, Intention and Behavior: An Introduction to Theory and Research*. Reading, MA: Addison-Wesley, 1975.

[22] B. H. Sheppard, J. Hartwick, and P. R. Warshaw, "The Theory of Reasoned Action: A Meta-Analysis of Past Research with Recommendations for Modifications and Future Research," *Journal of Consumer Research*, vol. 15, no. 3, pp. 325–343, 1988.

[23] C. Schmitz, "LimeSurvey Project Team." http://www.limesurvey.org, 2015.

[24] C. M. Ringle, S. Wende, and J. M. Becker, "SmartPLS 3." http://www.smartpls.com, 2015.

[25] J. Henseler, C. M. Ringle, and M. Sarstedt, "A new criterion for assessing discriminant validity in variance-based structural equation modeling," *Journal of the Academy of Marketing Science*, vol. 43, no. 1, pp. 115–135, 2015.

[26] N. K. Malhotra, S. S. Kim, and A. Patil, "Common Method Variance in IS Research: A Comparison of Alternative Approaches and a Reanalysis of Past Research," *Management Science*, vol. 52, no. 12, pp. 1865–1883, 2006.

[27] P. M. Podsakoff, S. B. MacKenzie, J. Y. Lee, and N. P. Podsakoff, "Common method biases in behavioral research: a critical review of the literature and recommended remedies.," *Journal of Applied Psychology*, vol. 88, no. 5, pp. 879–903, 2003.

[28] C. Blome and A. Paulraj, "Ethical Climate and Purchasing Social Responsibility: A Benevolence Focus," *Journal of Business Ethics*, vol. 116, no. 3, pp. 567–585, 2013.

[29] L. Rosen, K. Whaling, L. Carrier, N. Cheever, and J. Rokkum, "The Media and Technology Usage and Attitudes Scale: An empirical investigation," *Comput Human Behav.*, vol. 29, no. 6, pp. 2501–2511, 2013.

[30] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*. 1988.

[31] W. W. Chin, "The Partial Least Squares Approach to Structural Equation Modeling," in *Modern Methods for Business Research* (G. A. Marcoulides, ed.), pp. 295–336, Mahwah, NJ: Lawrence Erlbaum, 1998.

A. Distribution Channels of the Tor Online Survey

1. Mailinglists:
   (a) tor-talk[2]
   (b) liberationtech[3]
   (c) IFIP TC 11[4]
   (d) FOSAD[5]
   (e) GI PET[6]
   (f) GI FBSEC[7]
2. Twitter with #tor and #privacy
3. Boards:
   (a) reddit (sub-reddits: r/TOR, r/onions, r/privacy)
   (b) ubuntuusers.de
4. Tor Hidden Service Boards, Sections posted into:
   (a) Darknet Avengers[8], Off Topic
   (b) The Hub[9], Beginners
   (c) Onion Land[10], Off Topic
   (d) 8chan[11], /tech/
   (e) IntelExchange[12], Unverified Users
   (f) Code Green[13], Discussions
   (g) Changolia[14], overchan.random
   (h) Atlayo[15], Posting
5. Personal Announcements at Workshops

B. Questionnaire

The following items are measured with a seven-point Likert scale from "strongly disagree" to "strongly agree".

Trusting Beliefs (TB)

1. Online companies are trustworthy in handling information.
2. Online companies tell the truth and fulfill promises related to information provided by me.
3. I trust that online companies would keep my best interests in mind when dealing with information.
4. Online companies are in general predictable and consistent regarding the usage of information.
5. Online companies are always honest with customers when it comes to using the provided information.

Trusting Beliefs in Tor (TB_{Tor})

1. Tor is trustworthy.
2. Tor keeps promises and commitments.
3. I trust Tor because they keep my best interests in mind.

Risk Beliefs (RB)

1. In general, it would be risky to give information to online companies.
2. There would be high potential for loss associated with giving information to online firms.
3. There would be too much uncertainty associated with giving information to online firms.
4. Providing online firms with information would involve many unexpected problems.
5. I would feel safe giving information to online companies.

Awareness (AWA)

1. Companies seeking information online should disclose the way the data are collected, processed, and used.
2. A good consumer online privacy policy should have a clear and conspicuous disclosure.
3. It is very important to me that I am aware and knowledgeable about how my personal information will be used.

Collection (COLL)

1. It usually bothers me when online companies ask me for personal information.
2. When online companies ask me for personal information, I sometimes think twice before providing it.
3. It bothers me to give personal information to so many online companies.
4. Im concerned that online companies are collecting too much personal information about me.

Control (CONTROL)

1. Consumer online privacy is really a matter of consumers right to exercise control and autonomy over decisions about how their information is collected, used, and shared.
2. Consumer control of personal information lies at the heart of consumer privacy.
3. I believe that online privacy is invaded when control is lost or unwillingly reduced as a result of a marketing transaction.

Behavioral Intention (BI)

1. I intend to continue using Tor in the future.
2. I will always try to use Tor in my daily life.
3. I plan to continue to use Tor frequently.

Use Behavior (USE)

1. Please choose your usage frequency for Tor[16]
   - Never
   - Once a month
   - Several times a month
   - Once a week
   - Several times a week
   - Once a day
   - Several times a day
   - Once an hour
   - Several times an hour
   - All the time

[2] https://lists.torproject.org/cgi-bin/mailman/listinfo/tor-talk/
[3] https://mailman.stanford.edu/mailman/listinfo/liberationtech
[4] https://dlist.server.uni-frankfurt.de/mailman/listinfo/ifip-tc11
[5] http://www.sti.uniurb.it/events/fosad/
[6] http://mail.gi-fb-sicherheit.de/mailman/listinfo/pet
[7] http://mail.gi-fb-sicherheit.de/mailman/listinfo/fbsec
[8] http://avengersdutyk3xf.onion/
[9] http://thehub7xbw4dc5r2.onion
[10] http://onionlandbakyt3j.onion
[11] http://oxwugzccvk3dk6tj.onion
[12] http://rrcc5uuudhh4oz3c.onion
[13] http://pyl7a4ccwgpxm6rd.onion
[14] http://jewsdid.oniichanylo2tsi4.onion
[15] http://atlayofke5rqhsma.onion/
[16] The frequency scale is adapted from Rosen et al. [29].

C.9 How Privacy Concerns, Trust and Risk Beliefs and Privacy Literacy Influence Users' Intentions to Use Privacy-Enhancing Technologies - The Case of Tor

David Harborth and Sebastian Pape. How privacy concerns, trust and risk beliefs and privacy literacy influence users' intentions to use privacy-enhancing technologies - the case of Tor. *ACM SIGMIS Database: the DATABASE for Advances in Information Systems*, 51(1):51–69, January 2020. ISSN 0095-0033. doi: 10.1145/3380799.3380805. URL https://dl.acm.org/doi/abs/10.1145/3380799.3380805

HOW PRIVACY CONCERNS, TRUST AND RISK BELIEFS AND PRIVACY LITERACY INFLUENCE USERS' INTENTIONS TO USE PRIVACY-ENHANCING TECHNOLOGIES - THE CASE OF TOR

David Harborth
Chair of Mobile Business and Multilateral Security, Goethe University Frankfurt am Main

Dr. Sebastian Pape
Chair of Mobile Business and Multilateral Security, Goethe University Frankfurt am Main

Acknowledgments

This research was partly funded by the German Federal Ministry of Education and Research (BMBF) with grant number: 16KIS0371.

Abstract

Due to an increasing collection of personal data by internet companies and several data breaches, research related to privacy gained importance in the last years in the information systems domain. Privacy concerns can strongly influence users' decision to use a service. The Internet Users Information Privacy Concerns (IUIPC) construct is one operationalization to measure the impact of privacy concerns on the use of technologies. However, when applied to a privacy enhancing technology (PET) such as an anonymization service the original rationales do not hold anymore. In particular, an inverted impact of trusting and risk beliefs on behavioral intentions can be expected. We show that the IUIPC model needs to be adapted for the case of PETs. In addition, we extend the original causal model by including trusting beliefs in the anonymization service itself as well as a measure for privacy literacy. A survey among 124 users of the anonymization service Tor shows that trust in Tor has a statistically significant effect on the actual use behavior of the PET. In addition, the results indicate that privacy literacy has a negative impact on trusting beliefs in general and a positive effect on trust in Tor.

Keywords: Privacy Concerns; Tor; Privacy-Enhancing Technologies; Privacy Literacy; Technology Use

Introduction

"Surveillance is the business model of the internet. Everyone is under constant surveillance by many companies, ranging from social networks like Facebook to cellphone providers." (Mineo, 2017). Privacy and the related concerns have been discussed since the very beginning of computer sharing (David & Fano, 1965). Due to a raising economic interest in personal data during the last years (Bédard, 2016), privacy gains an increasing importance in individuals' everyday life. The majority of internet users has privacy concerns and feels a strong need to protect their privacy (Singh & Hill, 2003).

However, technologies which are able to protect users' privacy (PETs) are not widely adopted yet (Rossnagel, 2010). Among others, privacy concerns (Angst & Agarwal, 2009; Slyke, Johnson, Jiang, & Shim, 2006) and trust-risk-relationships (Harborth & Pape, 2018b, 2019; Smith, Dinev, & Xu, 2011) are assumed to have an important effect on the adoption of technologies. We argue that privacy concerns might have an important effect in the case of PETs, too. A popular model for measuring and explaining privacy concerns of online users is the model focusing on the Internet Users Information Privacy Concerns (IUIPC) construct by Malhotra, Kim, & Agarwal (2004). Their research involves a theoretical framework and an instrument for operationalizing privacy concerns, as well as a causal model for this construct including trust and risk beliefs about the online companies' data handling of personal information. The IUIPC construct has been used in various contexts, e.g. Internet of Things (Naeini et al., 2017), internet transactions (Heales, Cockcroft, & Trieu, 2017) and mobile apps (Raber & Krueger, 2017). Originally, the IUIPC instrument was applied to use cases for individuals' decisions to disclose personal information to service providers. However, for privacy enhancing technologies (PETs) the primary purpose is to help users to protect personal information when using regular internet services. As a consequence, it is necessary to reconsider the impact of trust and risk beliefs within IUIPC's causal model with respect to PETs. We expected this impact to be inverted and thus the trust model needs to be adapted for the investigation of PETs. In addition, trust in the PET itself is an important factor to consider. This is the case since Tor is used by a diverse group of people whose life might be endangered in case their identity is revealed (e.g. whistleblowers, opposition supporters, etc. (The Tor Project, 2018)). Besides users' concerns and trust, it is also important to consider the users' knowledge and capabilities. Users' attitudes often differ from the decisions they make ('privacy paradox') (Dienlin & Trepte, 2015). One way to explain the privacy paradox is that users balance between potential risks and benefits they gain from the service (privacy calculus) (Dinev & Hart, 2006). Another way to explain it is that users are concerned but lack knowledge to react in a way that would reflect their needs (Trepte et al., 2015).

Since we are surveying active users of Tor, both argumentations do not fit. In the former case, we have already explained that PETs are different than regular internet services since their primary goal is to protect the users' privacy. In the latter case, users have already installed the PET and use it. However, we still argue that it is important to consider the users' capabilities since users need a certain amount of knowledge in order to adequately evaluate the given level of privacy (Masur, Teutsch, & Trepte, 2017; Park, 2013). Thus, their knowledge might influence the users' trusting and risk beliefs in online companies and in particular the users' trusting beliefs in Tor. For that purpose, we measured the users' privacy literacy with the "Online Privacy Literacy Scale" (OPLIS) developed by Trepte et al. (2015).

To the best of our knowledge the OPLIS instrument in combination with the IUIPC construct has never been applied to a PET. Thus, we address the following research questions:

1. What influence have privacy concerns and associated trust and risk beliefs on the behavioral intention and actual use of Tor?
2. What influence does trust in Tor itself have on the behavioral intention and the actual use?
3. What influence does privacy literacy (measured with the OPLIS scale) have on trusting beliefs, risk beliefs and trusting beliefs in Tor?

For that purpose, we conducted an online survey with users of one of the most widely used anonymization services Tor (Tor has approximately 2,000,000 regular users) (The Tor Project, 2018). We collected 124 complete questionnaires out of 314 participants for the empirical analysis. Our results contribute to the understanding of users' perceptions about PETs and indicate how privacy concerns and trust and risk beliefs influence the use behavior of PETs.

The remainder of the paper is as follows: Section 2 introduces Tor and lists related work on PETs. In Section 3, we present research hypotheses and the data collection process. We assess the reliability and validity of our results in Section 4. In Section 5, we discuss the implications and limitations of our work and suggest future work. We conclude the article in Section 6.

Background and Related Work

Privacy-Enhancing Technologies (PETs) is an umbrella term for different privacy protecting technologies. PETs can be defined as a "coherent system of ICT measures that protects privacy [...] by eliminating or reducing personal data or by preventing unnecessary and/or undesired processing of personal data; all without losing the functionality of the data system" (Borking & Raab, 2001, p. 1).

Privacy Enhancing Technologies and Tor

In this paper, we investigate the privacy, trust and risk beliefs associated with PETs for the case of the anonymity service Tor. Tor is a free-to-use anonymity service that is based on the onion routing principle. The development of Tor started in 1995 in the Naval Research Lab (NRL). At that time the general idea was that one should be able to communicate over the Internet without revealing oneself to the other party (The Tor Project, 2018). Everybody can operate a server (relay) over which the encrypted traffic is routed. The routing occurs randomly over several different servers distributed world-wide. Tor aims to protect against an adversary who can observe or control some fraction of network traffic, but it does not protect against a global passive adversary, which means an adversary who can observe all network connections. Among the available PETs, Tor has one of the biggest user bases with approximately 2,000,000 active users (The Tor Project, 2018).

Related work on PETs considers mainly usability studies and to the best of our knowledge only two articles exist which focus on privacy concerns and related trust and risk beliefs of users of the PETs Tor and JonDonym (JonDos Gmbh, 2018). The two articles extend the IUIPC model by adding trust in the respective PET and find that the PET-specific trust (in Tor and JonDonym, respectively) has a large statistically significant positive effect on usage (Harborth & Pape, 2018b, 2019). Lee et al. (2017) assess the usability of the Tor Launcher and propose recommendations to overcome the found usability issues. Benenson, Girard, & Krontiris (2015) investigate acceptance factors for anonymous credentials. Among other things, they find that trust in the PET has no statistically significant impact on the intention to use the service. This result is relevant for our study since we hypothesize that trust in Tor has a positive effect on the actual use of the service (see Section 3.1). This hypothesis is supported by other research on technology acceptance factors of Tor which finds that trust in Tor is a highly relevant factor driving the use intention of the PET (Harborth & Pape, 2018a). Other research results indicate that trust in a PET has a positive effect on the willingness to pay money for this PET (Harborth, Cai, & Pape, 2019).

Privacy Concerns

A highly relevant study for our research is the one by Brecht et al. (2011), who investigate acceptance factors of anonymization services. Among other variables, they hypothesize a positive influence of privacy concerns on the intention to use such a service. Although they find a statistically significant effect, the effect is relatively small (effect size of 0.061) compared to other variables like perceived usefulness or internet privacy awareness. In contrast to our study, Brecht et al. (2011) use another operationalization of privacy concerns (Dinev & Hart, 2006) and they do not investigate it in the nomological network with trust and risk beliefs. However, it is highly relevant for constructs such as IUIPC and OPLIS to establish nomological validity (Straub, Boudreau, & Gefen, 2004). Therefore, we contribute to the theoretical discourse about privacy literacy and privacy concerns by including an operationalization of privacy literacy (OPLIS) and privacy concerns (IUIPC) in one nomological network with the trust-risk relationships

discussed before. We decided to use the operationalization for privacy concerns as in the original IUIPC paper (Malhotra et al., 2004). Thus, IUIPC is a second-order variable consisting of the constructs collection, control and awareness.

Online Privacy Literacy

Park (2013) defines online privacy literacy as a "principle to support, encourage, and empower users to undertake informed control of their digital identities". Trepte et al. (2015) give an exhaustive summary on the development of (online) privacy literacy. Several studies exist which aim to measure users' privacy literacy. Hoofnagle et al. (2010) ask users to answer whether five given statements about information handling of providers are true.

Brecht et al. (2012) find that users generally have a low knowledge about privacy issues on the Internet. They also find a negative correlation between a users' stated and their actual knowledge of privacy issues. Morrison (2013) investigates the same questions and asks ten objective questions and compares the results to three subjective questions (self-assessments). He finds that the users' self-assessment differs greatly from their objective knowledge about privacy. This cognitive bias where people mistakenly assess their cognitive ability as greater as than it is, is called Dunning-Kruger effect (Kruger & Dunning, 1999). As a consequence, users' statements on their knowledge about privacy cannot be trusted and other scales with users' self-assessments (cf. Park, 2013) are not further discussed here. Trepte et al. (2015) define online privacy literacy as "a combination of factual or declarative (knowing that) and procedural (knowing how) knowledge about online privacy" and implemented a scale based on "objective knowledge" to measure privacy literacy: the online privacy literacy scale (OPLIS). OPLIS consists of 20 questions divided into the following four knowledge groups:

4. practices of organizations, institutions and online service providers
5. technical aspects of data protection;
6. data protection law in Germany and Europe;
7. data protection strategies.

Since the constructs for data protection laws were specific for Germany and Europe and we surveyed Tor users worldwide, we needed to remove them (cf. Section 3.3).

Trepte and Masur (2017) apply a short version of OPLIS for a descriptive study. Joeckel and Dogruel (2019) investigate OPLIS, too. They find two correlations with the OPLIS score: a medium-sized with age (older participants know more about online privacy), and a weaker with privacy concerns (more privacy literate users were more concerned about their privacy). However, their correlation analysis does not offer any causality. Thus, it is unclear if more concerned users know more about privacy or users who know more are more concerned.

Methodology

We base our research on the Internet Users Information Privacy Concerns (IUIPC) model by Malhotra et al. (2004). The original research on this model investigates the role of users' information privacy concerns in the context of releasing personal information to a marketing service provider. Since we are focusing on the role of privacy concerns, trust and risk beliefs for the case of a PET (i.e. Tor), we adapt the original model according to the following logic. Originally, the service in question can be seen as the attacker (from a privacy point of view). If we apply the model to a service with the opposite goal, namely protecting the privacy of its users, certain relationships need to change. We will elaborate on the detailed changes in the next section. In addition, to this we extend the original model by trusting beliefs in the PET itself. We argue that the level of trust in a PET is a crucial factor determining the use decision.

For analyzing the cause-effect relationships between the latent (unobserved) variables, we use structural equation modelling (SEM). Since our research goal is to predict the target constructs behavioral intention and actual use behavior of Tor, we use partial least squares SEM (PLS-SEM) for our analysis (Hair, Hult, Ringle, & Sarstedt, 2017; Hair, Ringle, & Sarstedt, 2011) and not covariance-based SEM. In the following subsections, we discuss the hypotheses based on the IUIPC model (Malhotra et al., 2004), the questionnaire and the data collection process.

Research Hypotheses

The structural model contains several relationships between exogenous and endogenous variables (cf. Fig. 1). We develop our research hypotheses for these relationships along the hypotheses of the IUIPC model. IUIPC is operationalized as a second-order construct of the sub-constructs collection (COLL), awareness (AWA) and control

(CONTROL). Thus, the users' privacy concerns are determined by their concerns about "[...] individual-specific data possessed by others relative to the value of benefits receive" (Malhotra et al., 2004, p. 338), the control they have over their own data (i.e. possibilities to change or opt-out) and the "[...] degree to which a consumer is concerned about his/her awareness of organizational information privacy practices" (Malhotra et al., 2004, p. 339).

The effect of IUIPC on the behavioral intention is mediated by trusting beliefs and risk beliefs. Trusting beliefs are users' perceptions about the behavior of online firms to protect the users' personal information. In contrast, risk beliefs represent users' perception about losses associated with providing personal data to online firms (Malhotra et al., 2004). Thus, the higher the privacy concerns of a user, the lower are his or her trusting beliefs and the higher are his or her risk beliefs. In addition, a higher level of trust is assumed to decrease the risk beliefs. Thus, we hypothesize:

Hypothesis 1 (H1): Internet Users Information Privacy Concerns (IUIPC) have a negative effect on Trusting Beliefs (TB).

Hypothesis 2 (H2): Internet Users Information Privacy Concerns (IUIPC) have a positive effect on Risk Beliefs (RB).

Hypothesis 3 (H3): Trusting Beliefs (TB) have a negative effect on Risk Beliefs (RB).

Since we investigate the use of a specific PET, we extend the model by the trust in Tor itself with the adapted trust construct by Pavlou (2003). However, in order to protect their privacy, users with higher privacy concerns are assumed to rather trust the privacy-enhancing technology compared to online firms which process personal data. This is especially true, because we surveyed users of a PET which are assumed to take great care of their privacy. Therefore, we hypothesize:

Hypothesis 4 (H4): Internet Users Information Privacy Concerns (IUIPC) have a positive effect on the trusting beliefs in Tor (TB_{Tor}).

Privacy literacy is not a widely used concept in the information systems domain when investigating information privacy. Based on a representative selection of literature, Smith, Dinev, & Xu (2011) derive the "APCO" macro model summarizing related concepts and their relations to privacy concerns. However, no variable is or is related to privacy literacy. To the best of our knowledge, the construct we use in our analysis (OPLIS) is also not investigated in a nomological network with privacy concerns and outcome variables as behavioral intention or use. Therefore, we searched primarily for "privacy literacy" on Google Scholar as we argue that OPLIS operationalizes this concept. We find different applications and definitions of the concept "privacy literacy" in the literature. For example, online privacy literacy is defined as "[...] users' knowledge of privacy control tools (passive), and their actual application (active) to obscure the users' identity and protect his/her personal information on the internet" (Weinberger, Zhitomirsky-Geffet, & Bouhnik, 2017, p. 656). This definition is very focused on PETs and the research model is primarily looking at the antecedents of privacy literacy and the interrelations between the variables. For example, the results of the research indicate that online privacy concerns have a statistically significant positive effect on privacy literacy. However, due to a lack of a theoretical underlying we refrain from hypothesizing this relation since we argue that online privacy literacy is independent from IUIPC. Park (2013) investigates a closely related conceptualization of privacy literacy to OPLIS and the effect on corresponding behaviors in the digital sphere. The author finds that technical privacy knowledge, although very heterogenous amongst different demographic groups, has a positive correlation with users' ability to exert information control, i.e. decide about their personal information disclosure. OPLIS (Masur et al., 2017) was partially developed from skill items of the study by Park (2013) whereas the authors do not investigate OPLIS in a nomological network. "Social privacy literacy" as a sub-concept of privacy literacy for the case of social network is investigated in an article by Bartsch & Dienlin (2016). They find a positive effect of social privacy literacy on social privacy behavior. In summary, prior research suggests that privacy literacy might influence online behaviors positively in a way that individuals who are more literate behave in a more privacy-aware manner. However, privacy literacy is a context-independent variable comparable to IUIPC (Malhotra et al., 2004). Therefore, we argue that it behaves similar in its effects on the context-specific factors trusting beliefs, risk beliefs and trusting beliefs in Tor. As discussed before, previous research suggests that people with more privacy knowledge tend to be more aware about privacy threats (Bartsch & Dienlin, 2016), we argue that a higher level of online privacy literacy leads to less trust in online companies with respect to handling personal information. In contrast, risk beliefs will increase with a higher level of knowledge. Therefore, we hypothesize:

Hypothesis 5 (H5): Online Privacy Literacy (OPLIS) has a negative effect on Trusting Beliefs (TB).

Hypothesis 6 (H6): Online Privacy Literacy (OPLIS) has a positive effect on Risk Beliefs (RB).

The relationship of privacy literacy and trusting beliefs in Tor is not as clear as for hypotheses 5 and 6 because the OPLIS instrument does not contain any specific questions related to Tor. Thus, the assumption that trust in Tor is built upon the knowledge of certain specific features of Tor is difficult to make. However, since we asked active users of Tor, we argue that there is a certain level of trust in the service in place which is positively correlated with their relatively high knowledge related to privacy. We hypothesize this type of self-selection in hypothesis 7:

Hypothesis 7 (H7): Online Privacy Literacy (OPLIS) has a positive effect on the trusting beliefs in Tor (TB_{Tor}).

Trust is an important factor in the acceptance decision of users (Pavlou, 2003). McKnight, Carter, Thatcher, & Clay (2011) show that trust in a specific technology will positively affect individual's intention to explore the technology and to use more features of the technology in a postadoption context. Especially for the case of privacy protection, we assume that trust in the technology is a major factor for the intention to use the technology. For a further discussion on the concept of trust in a technology, we refer to Lankton, Mcknight, & Tripp (2015). We hypothesize that:

Hypothesis 8 (H8): Trusting beliefs in Tor (TB_{Tor}) have a positive effect on the behavioral intention to use Tor (BI).

It is logical that trusting beliefs have a positive effect and risk beliefs have a negative effect on releasing data and thus the intended behavior of using a regular service. However, for use behavior of a PET, we assume these effects reverse. The higher the trusting beliefs in online firms, the lower is the use frequency of Tor, since the protection of data becomes less important. Following this rationale, a higher degree of risk beliefs in data processing of online firms leads to a higher degree of use. Thus, we hypothesize that:

Hypothesis 9 (H9): Trusting beliefs (TB) have a negative effect on the behavioral intention to use Tor (BI).

Hypothesis 10 (H10): Risk beliefs (RB) have a positive effect on the behavioral intention to use Tor (BI).

Research on the relationship between behavioral intention and use behavior goes back to Fishbein & Ajzen (1975). Later research indicates a positive link between the two constructs (Sheppard, Hartwick, & Warshaw, 1988). Thus, we hypothesize that:

Hypothesis 11 (H11): The behavioral intention to use Tor (BI) has a positive effect on the actual use behavior (USE).

The resulting structural model is shown in Figure 1.

Insert Figure 1 About Here

Data Collection

The questionnaire constructs are adapted from the original IUIPC paper. The trust construct for trust in Tor is adapted from Pavlou (2003). Privacy literacy is operationalized with the online privacy literacy scale (OPLIS) (Masur et al., 2017). We conducted the study with German and English-speaking Tor users. Thus, we administered two questionnaires. All items for the German questionnaire had to be translated into German since all of the constructs are adapted from English literature. To ensure content validity of the translation, we followed a rigorous translation process. First, we translated the English questionnaire into German with the help of a certified translator (translators are standardized following the DIN EN 15038 norm). The German version was then given to a second independent certified translator who retranslated the questionnaire to English. This step was done to ensure the equivalence of the translation. Third, a group of five academic colleagues checked the two English versions with regard to this equivalence. All items were found to be equivalent. The items of the English version can be found in Appendix B.

Since we investigate the effect of privacy concerns, trust and risk beliefs on the use of Tor, we collected data of actual users. We installed the surveys on a university server and managed it with the survey software LimeSurvey (version 2.72.6) (Schmitz, 2015). The links to the English and German version were distributed over multiple channels on the internet. Although there are approximately 2,000,000 active users of the service, it was relatively difficult to gather the necessary number of complete answers for a valid and reliable quantitative analysis. Thus, to foster future research about Tor users, we provide an overview of every distribution channel in the Appendix A. In sum, 314 participants started the questionnaire (245 for the English version, 40 for the English version posted in hidden service forums and 29 for the German version). Of those 314 approached participants, 135 (105 for the English version, 13 for the English version posted in hidden service forums and 17 for the German version) filled

out the questionnaires completely. After deleting all sets from participants who answered a test question in the middle of the survey incorrectly, 124 usable data sets remained for the following analysis.

The demographic questions were not mandatory to fill out. This was done on purpose since we assumed that most of the participants are highly sensitive with respect to their personal data. Therefore, we had to resign from a discussion of the demographics in our research context. This decision is backed up by past research which does not find a statistically significant differences across gender, income groups, educational levels, or political affiliation in the desire to protect one's privacy (Singh & Hill, 2003).

Descriptive Statistics and OPLIS Adaption

The descriptive statistics for our quantitative analysis can be found in Table 1. The OPLIS value is calculated as a relative value (i.e. ratio of correctly answered questions divided by total number of questions).

As already mentioned in Section 2.3, we had to adapt the OPLIS score. The original questionnaire aimed at the German population. Thus, it contains questions about German and European data protection laws. Since our sample consists of Tor users possibly spread from all over the world, it does not make sense to ask them for German or even European law. As a consequence, we omitted the respective questions about national laws. This is straight forward since we consider the ratio of correctly answered questions. For a comparison with the reference group (cf. Figure 2), we extrapolate our results from 15 to 20 questions.

It can be seen that on average participants answered 78.78% of the questions correctly (with a median of 0.8). It can be seen that the participants are highly privacy-sensitive (median values for collection, awareness and control range from 6 to 7). This view is reinforced by a relatively low median value for trusting beliefs in online companies and an above neutral median value for risk beliefs. Trusting beliefs in Tor are relatively high with a median of 5.6667 indicating that most participants agree that they trust Tor. The descriptive statistics for the three covariates show that participants have on average almost 7 years of experience with Tor and almost 18 years of internet experience. This insight combined with the high privacy literacy implies that the sample is relatively knowledgeable and experienced compared to the general population of internet users. A median value of 4 indicates that participants perceive to be a victim of privacy breaches "occasionally".

Insert Table 1 About Here

The distribution of the cumulative relative frequency for correctly answered privacy literacy questions is illustrated in Figure 2.

As discussed in Section 2, we extrapolate our results for Tor users in order to make it comparable to results of a representative German sample of regular Internet users (Masur et al., 2017). The distribution graph clearly shows that the Tor users are more literate with respect to online privacy compared to regular German internet users. For example, 60% of the participants in our sample answered 12 out of 15 questions correctly (i.e. 80% correctly answered questions). In contrast, roughly 60% of the regular internet users in the reference group answered 12 out of 20 questions correctly (i.e. 60%).

Insert Figure 2 About Here

Results

We tested the model using SmartPLS version 3.2.7 (Ringle, Wende, & Becker, 2015). Before looking at the result of the structural model and discussing its implications, we discuss the measurement model, and check for the reliability and validity of our results. This is a precondition of being able to interpret the results of the structural model. Furthermore, it is recommended to report the computational settings. For the PLS algorithm, we choose the path weighting scheme with a maximum of 300 iterations and a stop criterion of 10−7. For the bootstrapping procedure, we use 5000 bootstrap subsamples and no sign changes as the method for handling sign changes during the iterations of the bootstrapping procedure.

Assessment of the Measurement Model

As the model is measured solely reflectively, we need to evaluate the internal consistency reliability, convergent validity and discriminant validity to assess the measurement model properly.

Internal consistency reliability (ICR) measurements indicate how well certain indicators of a construct measure the same latent phenomenon. Two standard approaches for assessing ICR are Cronbach's α and the composite reliability. The values of both measures should be between 0.7 and 0.95 for research that builds upon accepted models. Values of Cronbach's α are seen as a lower bound and values of the composite reliability as an upper bound of the assessment (Hair et al., 2017). Table 2 includes the ICR of the variables in the last two rows. It can be seen that all values for Cronbach's α are above the lower threshold of 0.7 except for RISK. However, for the composite reliability the value for RISK is higher than 0.7. Therefore, we argue that ICR is not a major issue for this variable. For all variables, no value is above 0.95. Values above that upper threshold indicate that the indicators measure the same dimension of the latent variable, which is not optimal with regard to the validity (Hair et al., 2017). In sum, ICR is established for our variables. Since IUIPC and USE are single-item constructs they have ICR values of 1.

Convergent validity determines the degree to which indicators of a certain reflective construct are explained by that construct. This is assessed by calculating the outer loadings of the indicators of the constructs (indicator reliability) and by looking at the average variance extracted (AVE). Loadings above 0.7 imply that the indicators have much in common, which is desirable for reflective measurement models. Table 2 shows the outer loadings in bold on the diagonal. All loadings were higher than 0.7, except for TRUST4 with a value of 0.289. Therefore, we dropped this item after an initial analysis. Convergent validity for the construct is assessed by the AVE. AVE is equal to the sum of the squared loadings divided by the number of indicators. A threshold of 0.5 is acceptable, indicating that the construct explains at least half of the variance of the indicators (Hair et al., 2017). The diagonal values of Table 3 present the AVE of our constructs. All values are well above 0.5, demonstrating convergent validity.

Insert Table 2 About Here

Discriminant validity measures the degree of uniqueness of a construct compared to other constructs. Comparable to the convergent validity assessment, two approaches are used for investigating discriminant validity. The first approach, assessing cross-loadings, is dealing with single indicators. All outer loadings of a certain construct should be larger than its cross-loadings with other constructs. Table 2 illustrates the cross-loadings as off-diagonal elements. All cross-loadings are smaller than the outer loadings, fulfilling the first assessment approach of discriminant validity. The second approach is on the construct level and compares the square root of the constructs' AVE with the correlations with other constructs. The square root of the AVE of a single construct should be larger than the correlation with other constructs (Fornell-Larcker criterion) (Hair et al., 2017). Table 3 contains the square root of the AVE on the diagonal in parentheses. All values are larger than the correlations with other constructs, indicating discriminant validity.

Insert Table 3 About Here

Since there are problems in determining the discriminant validity with both approaches, researchers propose the heterotrait-monotrait ratio (HTMT) for assessing discriminant validity as a superior approach (Henseler, Ringle, & Sarstedt, 2015). HTMT divides between-trait correlations by within-trait correlations, therefore providing a measure of what the true correlation of two constructs would be if the measurement is flawless (Hair et al., 2017). Values close to 1 for HTMT indicate a lack of discriminant validity. A conservative threshold is 0.85. Table 4 contains the values for HTMT and no value, except for the correlation between IUIPC and COLL (with 0.888), is above the threshold of 0.85. To assess if the HTMT statistics are significantly different from 1, we conducted a bootstrapping procedure with 5,000 subsamples to get the confidence interval in which the true HTMT value lies with a 95% chance. The HTMT measure requires that no confidence interval contains the value 1. The conducted analysis shows that this is the case, and thus discriminant validity is established for our model.

Insert Table 4 About Here

Common method bias (CMB) can occur if data is gathered with a self-reported survey at one point in time in one questionnaire (Malhotra, Kim, & Patil, 2006). Since this is the case in our research design, the need to test for CMB arises. An unrotated principal component factor analysis is performed with the software package STATA 14.0 to conduct the Harman's single-factor test to address the issue of CMB (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003). The assumptions of the test are that CMB is not an issue if there is no single factor that results from the factor analysis or that the first factor does not account for the majority of the total variance. The test shows that eight factors have eigenvalues larger than 1 which account for 72.86% of the total variance. The first factor explains 26.76% of the total variance. Based on the results of previous literature (Blome & Paulraj, 2013), we argue that CMB is not likely to be an issue in the data set.

Assessment and Results of the Structural Model

To assess the structural model, we evaluate possible collinearity problems, path coefficients, the level of adjusted R^2, the effect size f^2, the predictive relevance Q^2 and the effect size q^2. We address these evaluation steps to ensure the predictive power of the model with regard to the target constructs (Hair et al., 2017).

Collinearity is present if two predictor variables are highly correlated with each other. To address this issue, we assess the inner variance inflation factor (VIF). All VIFs above 5 indicate that collinearity between constructs is present. For our model, the highest VIF is 1.380. Thus, collinearity is apparently not an issue.

Figure 3 shows the results of the path estimations and the adjusted R^2-values of the endogenous variables BI and USE. The adjusted R^2 is 0.412 for BI and 0.055 for USE. Thus, our model explains 41.2% of the variance of BI and 5.5% of USE. There are different proposals for interpreting the size of this value. We choose to use the very conservative threshold proposed by Hair et al. (2011), where R^2 values are weak with values around 0.25, moderate with 0.50 and substantial with 0.75. Based on this classification, the R^2 value for BI is weak to moderate and for USE the value is very weak. For use behavior several participants answered that they never use Tor (21 participants answered never) although they stated to use the service several years (answers to the question: How many years are you using Tor? with a median of 6 years and an average of 6.87 years on a seven-point Likert scale). The correlation coefficient between the years of using Tor and the use frequency is very small, negative and statistically insignificant with -0.0222 and a p-value of 0.8066. These 21 answers massively bias the results for the relationship between behavioral intention and actual use behavior (the median value of use frequency is 5). However, we cannot explain why the participants answered like this. They either misunderstood the question, answered it intentionally like this to disguise their activity with Tor or found the scale for use behavior inappropriate. This might be due to the fact that the scale only contains once a month as the lowest use frequency besides never. It might be possible that these 21 users use Tor only a few times per year or that they used Tor some years ago and have not used it again since then. Therefore, they might have chosen never as an answer. However, we used an established scale to measure use behavior (Rosen, Whaling, Carrier, Cheever, & Rokkum, 2013), but recommend to consider this issue in future research studies with a similar context.

The path coefficients are presented on the arrows connecting the exogenous and endogenous constructs in Figure 3. Statistical significance is indicated by asterisks, ranging from three asterisks for p-values smaller than 0.01 to one asterisk for p-values smaller than 0.10. We chose this p-value range since p-values tend to be larger if the sample size is comparable small and we wanted to capture also significant effects above the 5% level. The p-value indicates the probability that a path estimate is incorrectly assumed to be significant. Thus, the lower the p-value, the higher the probability that the given relationship exists. The relevance of the path coefficients is shown by the relative size of the coefficient compared to the other explanatory variables (Hair et al., 2017).

Insert Figure 3 About Here

It can be seen that IUIPC has a relatively large statistically significant negative effect on trusting beliefs and a positive effect on risk beliefs. The effect of IUIPC on trusting beliefs in Tor is significant, positive and relatively weak compared to the other significant effects in the model. The construct trusting beliefs has a statistically significant medium-sized negative effect on risk beliefs. The effects of trusting beliefs and risk beliefs on behavioral intention are not statistically significant (for both $p \geq 0.10$). In contrast, the effect of trusting beliefs in Tor on behavioral intention is highly statistically significant, positive and large with 0.588. The second newly added construct OPLIS has a statistically significant negative impact on trusting beliefs in online companies and a positive effect on trusting beliefs in Tor. The effect on risk beliefs is not statistically significant.

The results for the covariates experience with Tor, internet experience and privacy victim experience are not depicted in Figure 3 due to clarity reasons. The results with the respective significance level are shown in Table 5. The results indicate that experience with Tor has no immediate effect on the five context-specific variables. Internet experience has a slightly significant negative effect on risk beliefs implying that experienced internet users tend to associate less risk with online companies handling their personal data. Personal privacy victim experiences exert statistically significant effects on trusting beliefs, trusting beliefs in Tor, behavioral intention as well as on the actual use behavior. The results indicate that a higher number of negative past experiences with privacy breaches lead to less trust in online companies. The same negative effect is in place for trust in Tor. Apparently, Tor users in our sample are well aware about the technical limitations of the PET with respect to protecting their anonymity. Therefore, they do not blindly assume that they are completely protected when using Tor. There is even the possibility that certain privacy breaches occurred while using a PET. Interestingly, at the same time there are positive effects on BI and USE. Thus, the overall result of the relations between privacy victim experiences, trust in Tor and behavioral intention and actual use behavior are rather ambiguous.

Insert Table 5 About Here

The f^2 effect size measures the impact of a construct on the endogenous variable by omitting it from the analysis and assessing the resulting change in the R^2 value. The values are assessed based on thresholds by Cohen (1988), who defines effects as small, medium and large for values of 0.02, 0.15 and 0.35, respectively. Table 6 shows the results of the f^2 evaluation. Values in italics indicate small effects, values in bold indicate medium effects and values in bold and italics indicate large effects. All other values have no substantial effect. The results correspond to those of the previous analysis of the path coefficients whereas trusting beliefs in Tor have a large effect on the behavioral intention.

The Q^2 measure indicates the out-of-sample predictive relevance of the structural model with regard to the endogenous latent variables based on a blindfolding procedure. We used an omission distance d=7. Recommended values for d are between five and ten (Hair et al., 2011). Furthermore, we report the Q^2 values of the cross-validated redundancy approach, since this approach is based on both the results of the measurement model as well as of the structural model. Values above 0 indicate that the model has the property of predictive relevance. In our case, the Q^2 value is equal to 0.306 for BI and 0.007 for USE. Since they are larger than zero, predictive relevance of the model is established.

The assessment of q^2 follows the same logic as the one of f^2. It is based on the Q^2 values of the endogenous variables and calculates the individual predictive power of the exogenous variables by omitting them and comparing the change in Q^2 (Hair et al., 2017). All individual values for q^2 are calculated with an omission distance d of seven. The results are shown in Table 6. The thresholds for the f^2 interpretation can be applied here, too. Values in italics indicate small effects, values in bold indicate medium effects and values in bold and italics indicate large effects. All other values have no substantial effect. As before, only the trusting beliefs in Tor have a medium-sized effect, implying the highest predictive power of all included exogenous variables. Risk beliefs have a small q^2 effect size.

Insert Table 6 About Here

Discussion

In this section, we interpret and summarize our findings of the statistical analysis, elaborate on limitations of our work and present future work opportunities.

Interpretation of the Results

Based on our results, all hypotheses except for H6, H9 and H10 can be confirmed (cf. Table 7). The results for H9 and H10 are surprising, considering that they are in contrast to the rationale explained in Section 3.1 and the results from previous literature (Malhotra et al., 2004). However, it must be said that when effect sizes are rather small it is possible that the relatively small sample size of 124 leads to a statistical non-significance. Thus, we cannot rule out that the effects of risk beliefs and trusting beliefs on behavioral intention would be significant with a larger sample size. Thus, only the degree of trust in the PET (Tor) has a direct significant effect on the intention to use the PET.

This result shows that a reputation of being trustworthy is crucial for a PET provider. The trusting beliefs in the PET itself are positively influenced by the users' information privacy concerns and their privacy literacy. Thus, the results imply that users with a higher level of privacy concerns and privacy literacy rather tend to trust a PET.

Insert Table 7 About Here

Hypothesis 6 cannot be confirmed, too. As for H9 and H10, the effect is not statistically significant. The hypotheses for the effects of privacy literacy on the two other context-specific factors trusting beliefs and risk beliefs can only be confirmed for the negative effect of OPLIS on trusting beliefs (H5). Thus, users who are more literate with respect to privacy tend to trust online companies less regarding the handling of their personal information. The effect of OPLIS on risk beliefs is not statistically significant.

Limitations

The limitations of the study primarily concern the sample composition and size. First, a larger sample would have been beneficial. However, in general, a sample of 124 participants is acceptable for our kind of statistical analysis and active users of a PET are hard to find for a relatively long online questionnaire. This is especially the case, if they do not have any financial rewards as in our study and if they are highly privacy sensitive which might repel them to disclose any kind of information (even if it is anonymous). Second, the combination of the results of the German and the English questionnaire can be a potential source of errors. German participants might have understood questions differently than the English participants. We argue that we achieved equivalence with regard to the meaning through conducting a thorough translation process, and therefore limiting this potential source of error to the largest extent possible. In addition, combining the data was necessary from a pragmatic point of view to get a sample size as large as possible for the statistical analysis. Third, we cannot rule out a non-response bias since especially in the privacy context people might not answer the questionnaire due to privacy concerns. Fourth, possible self-report biases (e.g. social desirability) might exist. We addressed these possible biases by gathering the data fully anonymized. As discussed earlier, we had issues with certain data sets of participants with regard to actual use behavior (cf. Section 4.2.). Although it might be more beneficial in certain settings to directly refer to actual use behavior as the sole target variable, we decided to include behavioral intention as an antecedent because of these issues. Lastly, our calculation of the OPLIS value is not based on all 20 questions of the original instrument since five questions are specific to law in the European Union. Thus, our results might not be comparable to the extent as we did in Figure 2. However, it is not possible to further break down the sample without the demographic information which we did not ask for mandatorily. Another limitation related to OPLIS concerns the validity of the instrument. OPLIS might have certain flaws since it is relatively new and not widely tested yet.

Future Work

Further work is required to investigate the specific determinants of use decisions for or against PETs and break down the interrelationships between the associated antecedents. In particular, it would be interesting to investigate the relationship between trusting beliefs in online companies and trust in the PET itself. A theoretical underlying would be required to include this relationship in such a research model. Furthermore, our work only investigates online literacy, especially online privacy literacy, in a specific context, i.e. with respect to the influence on specific variables and with respect to PETs. Thus, there is a lot of potential for future work to analyze this concept within different theories applied to different information systems. Interpreting privacy literacy as a kind of personal disposition might yield interesting results and might enable researchers to frame existing and new research questions based on another perspective. We also encourage building a more sophisticated model which not only includes privacy literacy but also closely related dimensions such as privacy awareness and the users' attitudes to investigate the users' intention and behavior.

Conclusions

In this paper, we contribute to the research on privacy-enhancing technologies and users' privacy by assessing the specific relationships between information privacy concerns, trusting beliefs in online firms and a privacy-enhancing technology (in our case Tor), risk beliefs associated with online firms' data processing, general privacy literacy and the actual use behavior of Tor. By adapting and extending the IUIPC model by Malhotra et al. (2004), we could show that several of the assumptions for regular online services do not hold for PETs. Furthermore, we contribute to the practical work on PETs, especially Tor, by providing insights into factors influencing use

intentions and behaviors of actual users. Trust in Tor is one of the major drivers of use intentions. Thus, companies or non-profits like 'The Tor Project' should focus on building a strong reputation and a trustful relationship with its users. We contribute to the literature on online literacy, by analyzing a relatively new instrument for measuring online privacy literacy (OPLIS) in two ways. First, our descriptive results of the OPLIS scores for Tor users indicate that they are more privacy literate than an average reference group of regular internet users (Masur et al., 2017). Second, we derived research hypotheses following the notion that online privacy literacy is similar to a personal disposition influencing the context-specific factors within the IUIPC model. Our results indicate that a higher level of online privacy literacy leads to less trust in online companies with respect to handling personal information. In contrast, more literate users tend to trust Tor to a larger extent. Thus, we argue that online privacy literacy is an important factor to consider when investigating relationships with privacy-related factors like concerns or risks.

References

Angst, C. M., & Agarwal, R. (2009). Adoption of Electronic Health Records in the Presence of Privacy Concerns: The Elaboration Likelihood Model and Individual Persuasion. MIS Quarterly, 33(2), 339–370.

Bartsch, M., & Dienlin, T. (2016). Control your Facebook: An analysis of online privacy literacy. Computers in Human Behavior, 56, 147–154. https://doi.org/10.1016/j.chb.2015.11.022

Bédard, M. (2016). The underestimated economic benefits of the internet. In Regulation series, The Montreal Economic Institute.

Benenson, Z., Girard, A., & Krontiris, I. (2015). User Acceptance Factors for Anonymous Credentials: An Empirical Investigation. 14th Annual Workshop on the Economics of Information Security (WEIS), 1–33.

Blome, C., & Paulraj, A. (2013). Ethical Climate and Purchasing Social Responsibility: A Benevolence Focus. Journal of Business Ethics, 116(3), 567–585. https://doi.org/10.1007/s10551-012-1481-5

Borking, J. J., & Raab, C. (2001). Laws, PETs and Other Technologies for Privacy Protection. Journal of Information, Law and Technology, 1, 1–14.

Brecht, F., Fabian, B., Kunz, S., & Mueller, S. (2011). Are You Willing to Wait Longer for Internet Privacy? In ECIS 2011 Proceedings. Retrieved from http://aisel.aisnet.org/ecis2011/236

Brecht, F., Fabian, B., Kunz, S., & Müller, S. (2012). Communication Anonymizers: Personality, Internet Privacy Literacy and Their Influence on Technology Acceptance. In ECIS 2012 Proceedings (pp. 1–13). Retrieved from http://aisel.aisnet.org/ecis2012/214

Cohen, J. (1988). Statistical Power Analysis for the Behavioral Sciences. HillsDale, NJ.

David, E. E., & Fano, R. M. (1965). Some thoughts about the social implications of accessible computing. In Proceedings 1965 Fall Joint Computer Conference.

Dienlin, T., & Trepte, S. (2015). Is the privacy paradox a relic of the past? An in-depth analysis of privacy attitudes and privacy behaviors. European Journal of Social Psychology, 45(3), 285–297. https://doi.org/10.1002/ejsp.2049

Dinev, T., & Hart, P. (2006). An extended privacy calculus model for e-commerce transactions. Information Systems Research, 17(1), 61–80. https://doi.org/10.1287/isre.1060.0080

Fishbein, M., & Ajzen, I. (1975). Belief, Attitude, Intention and Behavior: An Introduction to Theory and Research. Reading, MA: Addison-Wesley. https://doi.org/10.2307/2065853

Hair, J., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2017). A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM). SAGE Publications.

Hair, J., Ringle, C. M., & Sarstedt, M. (2011). PLS-SEM: Indeed a Silver Bullet. The Journal of Marketing Theory and Practice, 19(2), 139–152. https://doi.org/10.2753/MTP1069-6679190202

Harborth, D., Cai, X., & Pape, S. (2019). Why Do People Pay for Privacy-Enhancing Technologies? The Case of Tor and JonDonym. In G. Dhillon, F. Karlsson, K. Hedström, & A. Zúquete (Eds.), ICT Systems Security and Privacy Protection. SEC 2019. IFIP Advances in Information and Communication Technology, vol 562 (pp. 253–267). Springer, Cham. https://doi.org/https://doi.org/10.1007/978-3-030-22312-0_18

Harborth, D., & Pape, S. (2018a). Examining Technology Use Factors of Privacy-Enhancing Technologies: The Role of Perceived Anonymity and Trust. In Twenty-fourth Americas Conference on Information Systems. New Orleans, USA.

Harborth, D., & Pape, S. (2018b). JonDonym Users' Information Privacy Concerns. In L. Janczewski & M. Kutyłowski (Eds.), ICT Systems Security and Privacy Protection. SEC 2018. IFIP Advances in Information and Communication Technology, vol 529 (pp. 170–184). Poznan, Poland: Springer, Cham. https://doi.org/https://doi.org/10.1007/978-3-319-99828-2_13

Harborth, D., & Pape, S. (2019). How Privacy Concerns and Trust and Risk Beliefs Influence Users' Intentions to Use Privacy-Enhancing Technologies - The Case of Tor. In Hawaii International Conference on System

Sciences (HICSS) Proceedings (pp. 4851–4860). Hawaii, US.
Heales, J., Cockcroft, S., & Trieu, V.-H. (2017). The influence of privacy, trust, and national culture on internet transactions. In G. Meiselwitz (Ed.), Social Computing and Social Media. Human Behavior (pp. 159–176). Springer International Publishing.
Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. Journal of the Academy of Marketing Science, 43(1), 115–135. https://doi.org/10.1007/s11747-014-0403-8
Hoofnagle, C. J., King, J., Li, S., & Turow, J. (2010). How Different are Young Adults from Older Adults When it Comes to Information Privacy Attitudes and Policies? https://repository.upenn.edu/asc_papers/399. https://doi.org/10.2139/ssrn.1589864
Joeckel, S., & Dogruel, L. (2019). Default effects in app selection: German adolescents' tendency to adhere to privacy or social relatedness features in smartphone apps. Mobile Media & Communication, 1–20. https://doi.org/10.1177/2050157918819616
JonDos Gmbh. (2018). Official Homepage of JonDonym. Retrieved January 16, 2018, from https://www.anonym-surfen.de
Kruger, J., & Dunning, D. (1999). Unskilled and Unaware of It: How Difficulties in Recognizing One's Own Incompetence Lead to Inflated Self-Assessments. Journal of Personality and Social Psychology, 77(6), 1121–1134. https://doi.org/10.1037/0022-3514.77.6.1121
Lankton, N. K., Mcknight, D. H., & Tripp, J. (2015). Technology, Humanness, and Trust: Rethinking Trust in Technology. Journal of the Association for Information Systems, 16(10), 880–918.
Lee, L., Fifield, D., Malkin, N., Iyer, G., Egelman, S., & Wagner, D. (2017). A Usability Evaluation of Tor Launcher. Proceedings on Privacy Enhancing Technologies, (3), 90–109. https://doi.org/10.1515/popets-2017-0030
Malhotra, N. K., Kim, S. S., & Agarwal, J. (2004). Internet users' information privacy concerns (IUIPC): The construct, the scale, and a causal model. Information Systems Research, 15(4), 336–355. https://doi.org/10.1287/isre.1040.0032
Malhotra, N. K., Kim, S. S., & Patil, A. (2006). Common Method Variance in IS Research: A Comparison of Alternative Approaches and a Reanalysis of Past Research. Management Science, 52(12), 1865–1883. https://doi.org/10.1287/mnsc.1060.0597
Masur, P. K., Teutsch, D., & Trepte, S. (2017). Entwicklung und Validierung der Online-Privatheitskompetenzskala (OPLIS) [Development and validation of the Online Privacy Literacy Scale (OPLIS)]. Diagnostica, 63(4), 256–268. https://doi.org/10.1026/0012-1924/a000179
McKnight, D. H., Carter, M., Thatcher, J. B., & Clay, P. F. (2011). Trust in a Specific Technology: An Investigation of Its Components and Measures. ACM Transactions on Management Information Systems (TMIS), 2(2), 1–25. https://doi.org/10.1145/1985347.1985353
Mineo, L. (2017). On internet privacy, be very afraid (Interview with Bruce Schneier). Retrieved February 20, 2018, from https://news.harvard.edu/gazette/story/2017/08/when-it-comes-to-internet-privacy-be-very-afraid-analyst-suggests/
Morrison, B. (2013). Do We Know What We Think We Know? An Exploration of Online Social Network Users' Privacy Literacy. Workplace Review, April 2013.
Naeini, P. E., Bhagavatula, S., Habib, H., Degeling, M., Bauer, L., Cranor, L., & Sadeh, N. (2017). Privacy expectations and preferences in an iot world. In Symposium on Usable Privacy and Security (SOUPS).
Park, Y. J. (2013). Digital Literacy and Privacy Behavior Online. Communication Research, 40(2), 215–236. https://doi.org/10.1177/0093650211418338
Pavlou, P. A. (2003). Consumer Acceptance of Electronic Commerce: Integrating Trust and Risk with the Technology Acceptance Model. International Journal of Electronic Commerce, 7(3), 101–134. https://doi.org/10.1080/10864415.2003.11044275
Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: a critical review of the literature and recommended remedies. Journal of Applied Psychology, 88(5), 879–903. https://doi.org/10.1037/0021-9010.88.5.879
Raber, F., & Krueger, A. (2017). Towards understanding the influence of personality on mobile app permission settings. In IFIP Conference on Human-Computer Interaction (pp. 62–82).
Ringle, C. M., Wende, S., & Becker, J. M. (2015). SmartPLS 3. Boenningstedt: SmartPLS GmbH, http://www.smartpls.com. Retrieved from http://www.smartpls.com
Rosen, L. D., Whaling, K., Carrier, L. M., Cheever, N. A., & Rokkum, J. (2013). The Media and Technology Usage and Attitudes Scale: An empirical investigation. Comput Human Behav., 29(6), 2501–2511. https://doi.org/10.1016/j.pestbp.2011.02.012.Investigations
Rossnagel, H. (2010). The market failure of anonymity services. Lecture Notes in Computer Science (Incl.

Subseries Lecture Notes in AI and Lecture Notes in Bioinformatics), 6033 LNCS, 340–354. https://doi.org/10.1007/978-3-642-12368-9_28
Schmitz, C. (2015). LimeSurvey Project Team. Retrieved from http://www.limesurvey.org
Sheppard, B. H., Hartwick, J., & Warshaw, P. R. (1988). The Theory of Reasoned Action: A Meta-Analysis of Past Research with Recommendations for Modifications and Future Research. Journal of Consumer Research, 15(3), 325–343.
Singh, T., & Hill, M. E. (2003). Consumer privacy and the Internet in Europe: a view from Germany. Journal of Consumer Marketing, 20(7), 634–651.
Slyke, C. V., Johnson, R., Jiang, J., & Shim, J. T. (2006). Concern for Information Privacy and Online Consumer Purchasing. Journal of the Association for Information Systems, 7(6), 415–444.
Smith, H. J., Dinev, T., & Xu, H. (2011). Theory and Review Information Privacy Research: An Interdisciplinary Review. MIS Quarterly, 35(4), 989–1015.
Straub, D., Boudreau, M. C., & Gefen, D. (2004). Validation Guidelines for IS Positivist Research. Communications of the Association for Information Systems, 13, 380–427.
The Tor Project. (2018). Tor. Retrieved February 20, 2018, from https://www.torproject.org
Trepte, S., & Masur, P. K. (2017). Privacy attitudes, perceptions, and behaviors of the German population. Forum Privatheit Und Selbstbestimmung in Der Digitalen Welt.
Trepte, S., Teutsch, D., Masur, P. K., Eicher, C., Fischer, M., Hennhöfer, A., & Lind, F. (2015). Do People Know About Privacy and Data Protection Strategies? Towards the “Online Privacy Literacy Scale” (OPLIS). In S. Gutwirth, R. Leenes, & P. de Hert (Eds.), Reforming European Data Protection Law (Vol. 20). Springer Netherlands. https://doi.org/10.1007/978-94-017-9385-8
Weinberger, M., Zhitomirsky-Geffet, M., & Bouhnik, D. (2017). Factors affecting users’ online privacy literacy among students in Israel. Online Information Review, 41(5), 582–597. https://doi.org/10.1108/OIR-05-2016-0127

About the Authors

David Harborth holds a Master's degree in Management with specialization in Finance and Information Management and a Bachelor's degree in Business Administration and Economics with specialization in Finance and Accounting from Goethe University Frankfurt. He worked for three years in the consulting industry in the strategy and financial services sector. David Harborth started to work at the Chair of Mobile Business & Multilateral Security as a research and teaching assistant in December 2015. His major areas of research are the socio-economic and technical issues related to privacy in Augmented Reality (AR), as well as user perceptions and business models for privacy-enhancing technologies (PETs). His research has appeared in conferences such as *International Symposium on Mixed and Augmented Reality (ISMAR), Hawaii International Conference on System Sciences (HICSS), European Conference on Information Systems (ECIS), Americas Conference on Information Systems (AMCIS), Wirtschaftsinformatik (WI)* and *IFIP SEC*.

Dr. Sebastian Pape Sebastian Pape is a senior researcher working at the Chair of Mobile Business & Multilateral Security at Goethe University Frankfurt. He successfully completed diplomas in mathematics (Dipl.-Math.) and computer science (Dipl.-Inform.) at Darmstadt University of Technology and holds a doctoral degree (Dr. rer. nat.) from the University of Kassel. From 2005 to 2011, he worked as research and teaching assistant at the Database Group (lead by Prof. Dr. Lutz Wegner) of the Department of Electrical Engineering and Computer Science of the University of Kassel. From 2011 to 2015, he was a senior researcher and teaching assistant at the Software Engineering for Critical Systems Group (lead by Prof. Dr. Jan Jürjens) of the Department of Computer Science Department of TU Dortmund University. From October 2014 to January 2015, he also was a visiting researcher (of Prof. Dr. Fabio Massacci) at the security group of the Department of Information Engineering and Computer Science of University of Trento. From 2018 to 2019 he was standing in as a professor at the Chair of Information Systems of the Faculty of Business, Economics and Management Information Systems at Regensburg University.

Appendix A - Distribution Channels of the Tor Online Survey

1. *Mailinglists:*

(a) tor-talk (https://lists.torproject.org/cgi-bin/mailman/listinfo/tor-talk/)
(b) liberationtech (https://mailman.stanford.edu/mailman/listinfo/liberationtech)
(c) IFIP TC 11 (https://dlist.server.uni- frankfurt.de/mailman/listinfo/ifip-tc11)
(d) FOSAD (http://www.sti.uniurb.it/events/fosad/)
(e) GI PET (http://mail.gi-fb-sicherheit.de/mailman/listinfo/pet)
(f) GI FBSEC (http://mail.gi-fb-sicherheit.de/mailman/listinfo/fbsec)

2. *Twitter with #tor and #privacy*

3. *Boards:*

(a) reddit (sub-reddits: r/TOR, r/onions, r/privacy)
(b) ubuntuusers.de

4. *Tor Hidden Service Boards, Sections posted into:*

(a) Darknet Avengers, Off Topic (http://avengersdutyk3xf.onion/)
(b) The Hub, Beginners (http://thehub7xbw4dc5r2.onion)
(c) Onion Land, Off Topic (http://onionlandbakyt3j.onion)
(d) 8chan, /tech/ (http://oxwugzccvk3dk6tj.onion)
(e) IntelExchange, Unverified Users (http://rrcc5uuudhh4oz3c.onion)
(f) Code Green, Discussions (http://pyl7a4ccwgpxm6rd.onion)
(g) Changolia, overchan.random (http://jewsdid.oniichanylo2tsi4.onion)
(h) Atlayo, Posting (http://atlayofke5rqhsma.onion/)

5. *Personal Announcements at Workshops*

Appendix B - Questionnaire

The following items are measured with a seven-point Likert scale from "strongly disagree" to "strongly agree".

Trusting Beliefs (TB)

1. Online companies are trustworthy in handling information.
2. Online companies tell the truth and fulfill promises related to information provided by me.
3. I trust that online companies would keep my best interests in mind when dealing with information.
4. Online companies are in general predictable and consistent regarding the usage of information.
5. Online companies are always honest with customers when it comes to using the provided information.

Trusting Beliefs in Tor (TB_{Tor})

1. Tor is trustworthy.
2. Tor keeps promises and commitments.
3. I trust Tor because they keep my best interests in mind.

Risk Beliefs (RB)

1. In general, it would be risky to give information to online companies.
2. There would be high potential for loss associated with giving information to online firms.
3. There would be too much uncertainty associated with giving information to online firms.
4. Providing online firms with information would involve many unexpected problems.
5. I would feel safe giving information to online companies. (reverse-scored item)

Awareness (AWA)

1. Companies seeking information online should disclose the way the data are collected, processed, and used.
2. A good consumer online privacy policy should have a clear and conspicuous disclosure.
3. It is very important to me that I am aware and knowledgeable about how my personal information will be used.

Collection (COLL)

1. It usually bothers me when online companies ask me for personal information.
2. When online companies ask me for personal information, I sometimes think twice before providing it.
3. It bothers me to give personal information to so many online companies.
4. I'm concerned that online companies are collecting too much personal information about me.

Control (CONTROL)

1. Consumer online privacy is really a matter of consumers right to exercise control and autonomy over decisions about how their information is collected, used, and shared.
2. Consumer control of personal information lies at the heart of consumer privacy.
3. I believe that online privacy is invaded when control is lost or unwillingly reduced as a result of a marketing transaction.

Behavioral Intention (BI)
1. I intend to continue using Tor in the future.
2. I will always try to use Tor in my daily life.
3. I plan to continue to use Tor frequently.

Use Behavior (USE)
Please choose your usage frequency for Tor (the frequency scale is adapted from Rosen et al. (2013)):
1. Never
2. Once a month
3. Several times a month
4. Once a week
5. Several times a week
6. Once a day
7. Several times a day
8. Once an hour
9. Several times an hour
10. All the time

Internet Experience (in years)
1. How many years of experience do you have with computers?

Answer options range from 0 years to “more than 20 years”.

Experience with Tor (in years)
1. How many years are you using Tor?

Answer options range from 0 years to “more than 20 years”.

Privacy Victim Experience
1. How frequently have you personally been the victim of what you felt was an improper invasion of privacy?

Item measured with a seven-point frequency scale (“Never”, “Very infrequently”, “Infrequently”, “Occasionally”, “Sometimes”, “Frequently”, “Very frequently”).

Online Privacy Literacy Scale (OPLIS)
Part 1: Knowledge about institutional practices
1. The National Security Agency (NSA) accesses only public user data, which are visible for anyone. (True/**false**/don’t know)
2. Social network site operators (e.g. Facebook) also collect and process information about non-users of the social network site. (**True**/false/don’t know)
3. User data that are collected by social network site operators (e.g. Facebook) are deleted after five years. (True/**false**/don’t know)
4. Companies combine users’ data traces collected from different websites to create user profiles. (**True**/false/don’t know)
5. E-mails are commonly passed over several computers before they reach the actual receiver. (**True**/false/don’t know)

Part 2: Knowledge about technical aspects of data protection (correct answers randomized)
1. What does the term “browsing history” stand for? In the browsing history...
   A. ...the URLs of visited websites are stored.
   B. ...cookies from visited websites are stored.
   C. ...potentially infected websites are stored separately.
   D. ...different information about the user are stored, depending on the browser type.
2. What is a “cookie”?
   A. A text file that enables websites to recognize a user when revisiting.
   B. A program to disable data collection from online operators.
   C. A computer virus that can be transferred after connecting to a website.
   D. A browser plugin that ensures safe online surfing.

3. What does the term "cache" mean?
   A. A buffer memory that accelerates surfing on the Internet.
   B. A program that specifically collects information about an Internet user and passes them on to third parties.
   C. A program, that copies data on an external hard drive to protect against data theft.
   D. A browser plugin that encrypts data transfer when surfing online.
4. What is a "trojan"? A trojan is a computer program, that...
   A. ...is disguised as a useful application, but fulfills another function in the background.
   B. ...protects a computer from viruses and other malware.
   C. ... was developed for fun an d has no specific function.
   D. ... caused damage as computer virus in the 90ies but doesn't exist anymore.
5. What is a "firewall"?
   A. A fallback system that will protect the computer from unwanted web attacks.
   B. An outdated protection program against computer viruses.
   C. A browser plugin that ensures safe online surfing.
   D. A new technical development that prevents data loss in case of a short circuit.

Part 3: Knowledge about data protection strategies

1. Tracking of one's own internet is made more difficult if one deletes browser information (e.g. cookies, cache, browser history) regularly. (**True**/false/don't know)
2. Surfing in the private browsing mode can prevent the reconstruction of your surfing behavior, because no browser information is stored. (**True**/false/don't know)
3. Using false names or pseudonyms can make it difficult to identify someone on the Internet. (**True**/false/don't know)
4. Even though It-experts can crack difficult passwords, it is more sensible to use a combination of letters, numbers and signs as passwords than words, names or simple combinations of numbers. (**True**/false/don't know)
5. In order to prevent the access to personal data, one should use various passwords and user names for different online applications and change them frequently. (**True**/false/don't know)

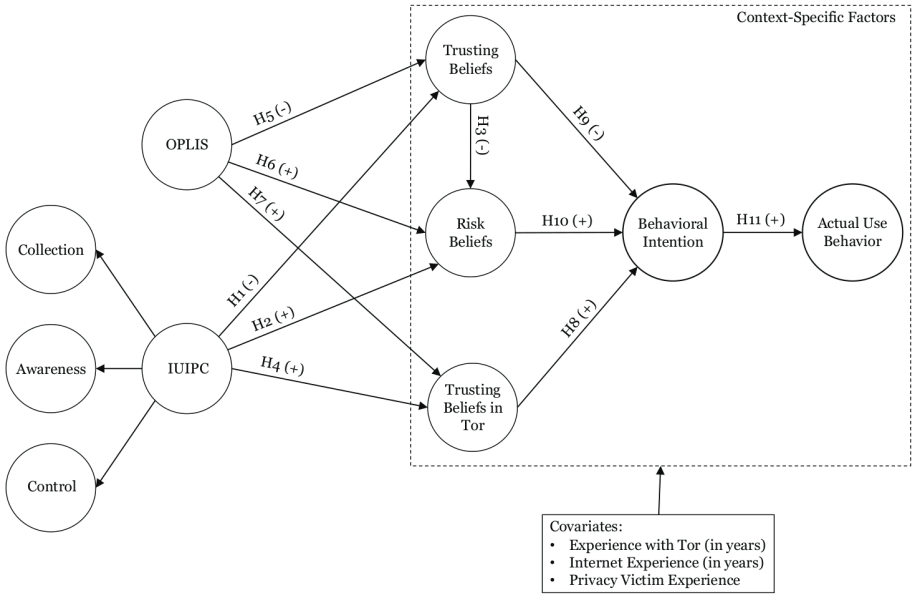


Figure 1. Research Model

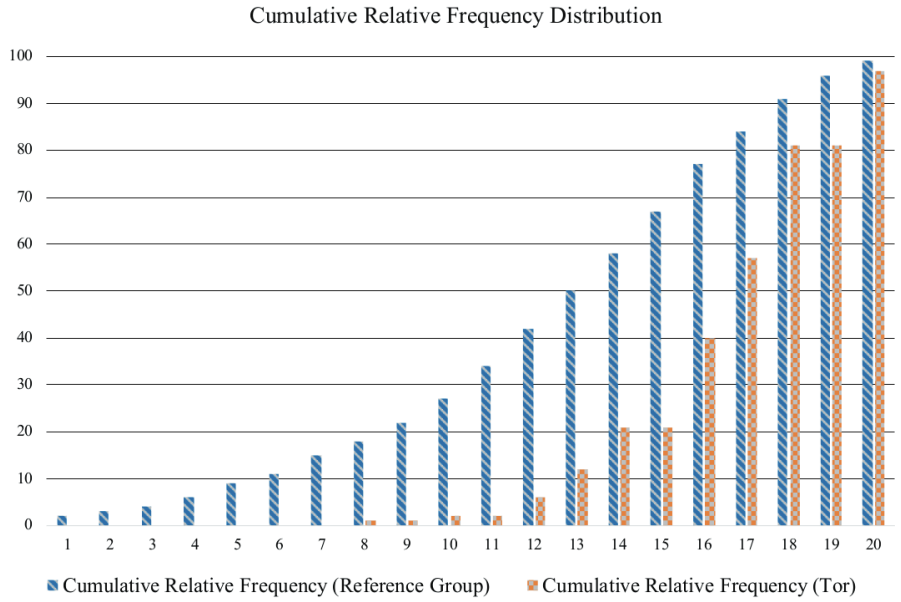


Figure 2. Differences in the Distributions between the Cumulative Relative Frequency of Correctly Answered OPLIS Questions between the Reference Group (Masur et al., 2017) and the Tor Users in our Sample

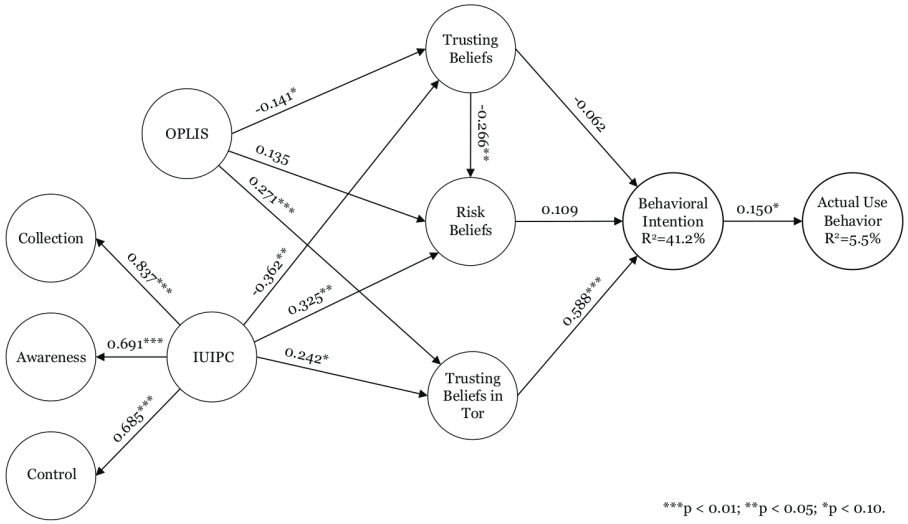


Figure 3. Path Estimates and Adjusted R^2 Values of the Structural Model

Table 1. Descriptive Statistics for the Used Variables (cf. Appendix A2 for Measurement Scales of the Constructs)

Statistics / Variable	Mean	Median	Min.	Max.	Std. Dev.
OPLIS (relative)	0.7876	0.8	0.3333	1	0.1259
Collection	6.3810	6.5	4	7	0.7053
Awareness	6.5457	7	1	7	0.7500
Control	5.9435	6	1	7	1.0038
Trusting Beliefs	2.2694	2.2	1	5.8	0.9429
Risk Beliefs	5.3242	5.5	1.6	7	1.1048
Trusting Beliefs in Tor	5.3548	5.6667	1	7	1.1892
Behavioral Intention	5.7043	6	1	7	1.2971
Actual Use Behavior	4.0726	5	0	9	2.6692
Experience with Tor	6.8710	6	0	20	4.6416
Internet Experience	17.7984	21	2	21	5.0429
Privacy Victim Experience	4.2742	4	1	7	1.6297

Table 2. Loadings and Cross-Loadings of the Reflective Items and Internal Consistency Reliability

Construct	AWA	Control	COLL	RB	TB	$Trust_{Tor}$	BI
AWA1	**0.911**	0.234	0.302	0.222	-0.136	0.066	0.201
AWA2	**0.923**	0.230	0.219	0.136	-0.153	0.072	0.197
AWA3	**0.891**	0.323	0.315	0.220	-0.102	0.066	0.249
CONTROL1	0.095	**0.825**	0.271	0.107	-0.163	0.137	0.214
CONTROL2	0.405	**0.821**	0.226	0.245	-0.149	0.132	0.237
CONTROL3	0.174	**0.756**	0.438	0.214	-0.340	0.098	0.098
COLL1	0.264	0.358	**0.888**	0.546	-0.462	0.176	0.301
COLL2	0.206	0.332	**0.812**	0.204	-0.337	0.232	0.374
COLL3	0.292	0.359	**0.906**	0.443	-0.442	0.272	0.375
COLL4	0.304	0.309	**0.850**	0.466	-0.399	0.182	0.317
RISK1	0.196	0.200	0.487	**0.879**	-0.446	0.217	0.258
RISK2	0.170	0.160	0.326	**0.832**	-0.292	0.156	0.233
RISK3	0.155	0.252	0.364	**0.861**	-0.346	0.233	0.221
RISK4	0.245	0.231	0.374	**0.826**	-0.255	0.257	0.327
RISK5	-0.105	-0.145	-0.427	**-0.700**	0.396	-0.003	-0.144
TRUST1	-0.149	-0.261	-0.455	-0.417	**0.894**	-0.097	-0.265
TRUST2	-0.118	-0.186	-0.410	-0.376	**0.890**	-0.033	-0.195
TRUST3	-0.107	-0.339	-0.397	-0.396	**0.768**	-0.131	-0.153
TRUST5	-0.069	-0.009	-0.219	-0.069	**0.682**	-0.109	-0.166
$TRUST_{Tor}1$	0.064	0.149	0.257	0.159	-0.091	**0.880**	0.559
$TRUST_{Tor}2$	0.077	0.121	0.236	0.244	-0.124	**0.924**	0.552
$TRUST_{Tor}3$	0.059	0.138	0.169	0.179	-0.078	**0.883**	0.486
BI1	0.236	0.240	0.355	0.228	-0.252	0.586	**0.858**
BI2	0.262	0.202	0.322	0.318	-0.149	0.465	**0.864**
BI3	0.143	0.158	0.363	0.233	-0.231	0.522	**0.926**
Cronbach's α	0.894	0.722	0.887	0.567	0.831	0.877	0.859
Comp. Reliability	0.934	0.843	0.922	0.817	0.885	0.924	0.914

Table 3. Discriminant Validity with AVEs and Construct Correlations

Constructs (AVE)	AWA	BI	COLL	Control	IUIPC	OPLIS	RB	TB	$Trust_{Tor}$	USE
AWA (0.825)	0.908									
BI (0.780)	0.239	0.883								
COLL (0.748)	0.309	0.394	0.865							
Control (0.642)	0.291	0.226	0.393	0.801						
IUIPC (1.000)	0.691	0.403	0.837	0.685	1.000					
OPLIS (1.000)	-0.071	0.143	0.111	0.110	0.071	1.000				
RB (0.675)	0.214	0.290	0.485	0.243	0.450	0.198	0.822			
TB (0.662)	-0.142	-0.242	-0.476	-0.276	-0.426	-0.155	-0.426	0.813		
$Trust_{Tor}$ (0.803)	0.075	0.597	0.249	0.152	0.226	0.300	0.217	-0.110	0.896	
USE (1.000)	-0.128	0.177	0.073	0.008	-0.009	0.006	0.010	-0.058	-0.026	1.000

Note: AVEs in parentheses in the first column. Values for √AVE are shown on the diagonal and construct correlations are off-diagonal elements.

Table 4. Heterotrait-Monotrait Ratio (HTMT)

Constructs	AWA	BI	COLL	Control	IUIPC	OPLIS	RB	TB	$Trust_{Tor}$
BI	0.274								
COLL	0.343	0.452							
Control	0.346	0.290	0.486						
IUIPC	0.728	0.436	0.888	0.798					
OPLIS	0.075	0.155	0.119	0.127	0.071				
RB	0.238	0.337	0.541	0.294	0.478	0.212			
TB	0.159	0.278	0.528	0.336	0.439	0.171	0.449		
$Trust_{Tor}$	0.084	0.681	0.280	0.192	0.240	0.318	0.244	0.131	
USE	0.138	0.186	0.077	0.060	0.009	0.006	0.021	0.058	0.029

Table 5. Covariate Results (Significance Levels: ***p < 0.01; **p < 0.05; *p < 0.10)

Context-specific factors / Covariate	TB	RB	TB_{Tor}	BI	USE
Experience with Tor	-0.047	-0.008	-0.012	0.092	-0.074
Internet experience	-0.001	-0.139*	0.003	0.016	0.065
Privacy victim experience	-0.245**	0.011	-0.163*	0.196**	0.225**

Table 6. f^2 and q^2 Effect Size Assessment Values

Variables	**f^2**	**q^2**
Endogenous / Exogenous	BI	BI
TB	0.005	0.000
RB	0.016	*0.072*
TB_{Tor}	***0.567***	**0.334**

Table 7. Summary of the Results

	Hypothesis	Result
H1	Internet Users Information Privacy Concerns (IUIPC) have a negative effect on Trusting Beliefs (TB)	√
H2	Internet Users Information Privacy Concerns (IUIPC) have a positive effect on Risk Beliefs (RB)	√
H3	Trusting Beliefs (TB) have a negative effect on Risk Beliefs (RB)	√
H4	Internet Users Information Privacy Concerns (IUIPC) have a positive effect on the trusting beliefs in Tor (TB_{Tor})	√
H5	Online Privacy Literacy (OPLIS) has a negative effect on Trusting Beliefs (TB)	√
H6	Online Privacy Literacy (OPLIS) has a positive effect on Risk Beliefs (RB)	×
H7	Online Privacy Literacy (OPLIS) has a positive effect on the trusting beliefs in Tor (TB_{Tor})	√
H8	Trusting beliefs in Tor (TB_{Tor}) have a positive effect on the behavioral intention to use Tor (BI)	√
H9	Trusting beliefs (TB) have a negative effect on the behavioral intention to use Tor (BI)	×
H10	Risk beliefs (RB) have a positive effect on the behavioral intention to use Tor (BI)	×
H11	The behavioral intention to use Tor (BI) has a positive effect on the actual use behavior (USE)	√

C.10 Explaining the Technology Use Behavior of Privacy-Enhancing Technologies: The Case of Tor and JonDonym

David Harborth, Sebastian Pape, and Kai Rannenberg. Explaining the technology use behavior of privacy-enhancing technologies: The case of Tor and JonDonym. *Proceedings on Privacy Enhancing Technologies (PoPETs)*, 2020(2):111–128, 2020. doi: 10.2478/popets-2020-0020. URL https://content.sciendo.com/view/journals/popets/2020/2/article-p111.xml

sciendo

Proceedings on Privacy Enhancing Technologies ; 2020 (2):111–128

David Harborth*, Sebastian Pape, and Kai Rannenberg

Explaining the Technology Use Behavior of Privacy-Enhancing Technologies: The Case of Tor and JonDonym

Abstract: Today's environment of data-driven business models relies heavily on collecting as much personal data as possible. Besides being protected by governmental regulation, internet users can also try to protect their privacy on an individual basis. One of the most famous ways to accomplish this, is to use privacy-enhancing technologies (PETs). However, the number of users is particularly important for the anonymity set of the service. The more users use the service, the more difficult it will be to trace an individual user. There is a lot of research determining the technical properties of PETs like Tor or JonDonym, but the use behavior of the users is rarely considered, although it is a decisive factor for the acceptance of a PET. Therefore, it is an important driver for increasing the user base.

We undertake a first step towards understanding the use behavior of PETs employing a mixed-method approach. We conducted an online survey with 265 users of the anonymity services Tor and JonDonym (124 users of Tor and 141 users of JonDonym). We use the technology acceptance model as a theoretical starting point and extend it with the constructs *perceived anonymity* and *trust in the service* in order to take account for the specific nature of PETs. Our model explains almost half of the variance of the *behavioral intention* to use the two PETs. The results indicate that both newly added variables are highly relevant factors in the path model. We augment these insights with a qualitative analysis of answers to open questions about the users' concerns, the circumstances under which they would pay money and choose a paid premium tariff (only for JonDonym), features they would like to have and why they would or would not recommend Tor/JonDonym. Thereby, we provide additional insights about the users' attitudes and perceptions of the services and propose new use factors not covered by our model for future research.

Keywords: Privacy-Enhancing Technologies, Tor, JonDonym, user study, technology acceptance

DOI 10.2478/popets-2020-0020
Received 2019-08-31; revised 2019-12-15; accepted 2019-12-16.

***Corresponding Author: David Harborth:** Goethe University Frankfurt, Germany, E-mail: david.harborth@m-chair.de

Sebastian Pape: Goethe University Frankfurt, Germany, E-mail: sebastian.pape@m-chair.de

Kai Rannenberg: Goethe University Frankfurt, Germany, E-mail: kai.rannenberg@m-chair.de

1 Introduction

Perry Barlow [6] states: "The internet is the most liberating tool for humanity ever invented, and also the best for surveillance. It's not one or the other. It's both." One of the reasons for surveilling users is a rising economic interest in the internet [7]. However, users who have privacy concerns and feel a strong need to protect their privacy are not helpless, they can make use of privacy-enhancing technologies (PETs). PETs allow users to improve their privacy by eliminating or minimizing personal data disclosure to prevent unnecessary or unwanted processing of personal data [58]. Examples of PETs include services which allow anonymous communication, such as Tor [56] or JonDonym [35].

There has been lots of research on Tor and JonDonym [43, 50], but the large majority of it is of technical nature and does not consider the user. However, the number of users is crucial for this kind of services. Besides the economic point of view which suggests that more users allow a more cost-efficient way to run those services, the quality of the offered service is depending on the number of users since an increasing number of (active) users also increases the anonymity set. The anonymity set is the set of all possible subjects who might be related to an action [46], thus a larger anonymity set may make it more difficult for an attacker to identify the sender or receiver of a message [2]. As a consequence, it's crucial to learn about the users' intention to use a PET and investigate the factors it depends on. Thus, our research is in line with related work on the obstacles of using secure communication tools [1] with the recommendation to "understand the target population" and research suggesting zero-effort privacy [28, 32] by improving the usability of the service.

In this paper, we investigate how the users' perceived anonymity and their trust in the service influence the intention to use PETs. Privacy protection is usually not the primary goal of the users, but only their secondary goal [17]. The user's

aims become more indistinct if the PET is integrated in the regular service (e.g. anonymous credentials [8]). In contrast to PETs integrated in services, "standalone" PETs are not integrated into a specific service and can be used for several purposes. Thus, examining standalone PETs allows us to focus on the usefulness of the PET with regard to privacy protection and avoids interference with other goals of the user. Therefore, we conducted a survey of the users of the (standalone) anonymity services Tor and JonDonym. The similarities and differences of the two considered PETs are sketched in the next section.

To determine the use factors of Tor and JonDonym, we extend the classical technology acceptance factors by Davis [18, 19] with relevant factors for the specific nature of PETs. We focus on *perceived anonymity* and *trust* because the perception about anonymity is a key variable for users to decide whether to use a such services or not. This perception is closely related to the trust which users might have in services. For example, there are vivid discussions with people claiming that Tor is essentially a big honeypot controlled by the US government. Opposing voices argue that anonymity is never achievable to 100% and that Tor is among the better solutions we have for certain scenarios (e.g. see a recent discussion which developed after a Twitter tweet by Edward Snowden on Tails [57]).

Since most users do not base their decisions on any kind of formal (technical or mathematical) anonymity measurement, we decided to measure the perceived anonymity. The resulting research question is:

RQ1: Does perceived anonymity influence the behavioral intention to use a PET?

However, *perceived anonymity* is a subjective perception of each user. Since we assume, that most users will not dig into mathematical proofs of the assured anonymity or challenge the implementation of the service provider, we conclude that it is important to also consider the *trust in the service provider and the service* itself:

RQ2: Does trust in the PET influence the behavioral intention to use it?

We further refine the two research questions and in particular the relation between *perceived anonymity*, *trust in the service* (Tor/JonDonym), *perceived usefulness*, *perceived ease of use*, *behavioral intention* and *actual use behavior* in Section 3. Consequently, the question arises whether the relationships between the variables of the model differ for the two PETs. We address this question by comparing the results based on a multigroup analysis. To augment and generalize the findings, we also asked users open questions about their concerns, their willingness to donate to Tor or use JonDonym's (paid) premium service, features they would like to have and why they would or would not recommend Tor/JonDonym.

The remainder of the paper is structured as follows: Section 2 briefly introduces the anonymization services Tor and JonDonym, provides information on the technology acceptance model and lists related work on PETs and technology acceptance. In Section 3, we present the research hypotheses, describe the questionnaire and the data collection process. We assess the quality of our quantitative empirical results with regard to reliability and validity in Section 4. We present the results for the research model for PETs and the multigroup analysis to compare Tor and JonDonym in Section 5 and for the qualitative analysis of the open questions in Section 6. In Section 7, we discuss the implications of the results, elaborate on limitations of our work and present possible future work. Section 8 concludes the paper with a summary of the findings.

2 Theoretical Background

Privacy-Enhancing Technologies (PETs) is an umbrella term for different privacy protecting technologies. Borking and Raab define PETs as a "coherent system of ICT measures that protects privacy [. . .] by eliminating or reducing personal data or by preventing unnecessary and/or undesired processing of personal data; all without losing the functionality of the data system" [10, p.1].

PETs have a property that is not characteristic for many other technology types. Privacy protection is usually not the primary goal of the users, but only their secondary goal [17]. It is important to understand that in many cases PET users make use of the PET while they pursue another goal like browsing the internet or using instant messengers. These aims become more indistinct if the PET is integrated in the regular service (e.g. anonymous credentials [8]). In contrast to PETs integrated in services, standalone PETs (e.g. overlay networks like Tor [56] or JonDonym [35]) are not integrated into a specific service and can be used for several purposes.

In this paper, we investigate the role of *perceived anonymity* and *trust* in the context of a technology acceptance model for the case of standalone PETs, namely the anonymity services Tor and JonDonym.

2.1 Tor and JonDonym

Tor and JonDonym are low latency anonymity services which redirect packets in a certain way in order to hide metadata (the sender's and optionally – in case of a hidden service – the receiver's internet protocol (ip) address) from passive network observers. In contrast to anonymity services with higher latency such as anonymous remailers low latency anonymity services can be used for interactive services such as messen-

gers. Due to network overheads this still leads to increased latency which was evaluated by Fabian et al. [21] who found associated usability issues when using Tor.

Technically, Tor – the onion router – is an overlay network where the users' traffic is encrypted and directed over several different servers (relays). The Tor client gets a file with a list of relays and follows a certain algorithm to select some relays for a circuit. The aim of the algorithm is to avoid to have two relays in one circuit which are run by the same entity. Selected routes through the circuit should be difficult for an adversary to observe. Consequently, unpredictable routes through the Tor network are chosen. The relays where the traffic leaves the Tor network are called "exit nodes" and for an external service the traffic seems to originate from those. JonDonym is based on user selectable mix cascades (a group of anonymization proxies), with two or three mix servers in one cascade. For mix networks route unpredictability is not important so within one cascade always the same sequence of mix servers is used. Thus, for an external service the traffic seems to originate from the last mix server in the cascade. As a consequence, other usability issues may arise when websites face some abusive traffic from the anonymity services [53] and decide to restrict access for users of the anonymity service. Restrictions range from outright rejection to limiting the users' access to a subset of the service's functionality or imposing hurdles such as CAPTCHA-solving [36] and for the user it appears that the website is not function properly.

Tor offers an adapted browser including the Tor client for using the Tor network, the "Tor Browser". Similarly, the "JonDoBrowser" includes the JonDo client for using the JonDonym network.

Although the specific technical functioning differ, JonDonym and Tor are highly comparable with respect to the general technical structure and the use cases. However, the entities who operate the PETs are different. Tor is operated by a non-profit organization with thousands of voluntarily operated servers (relays) over which the encrypted traffic is directed. Tor is free to use with the option that users can donate to the Tor project. The actual number of users is estimated with approximately 2,000,000 daily users by the Tor Project [56]. However, a recent study using another measurement technique found 8,000,000 daily users [42]. JonDonym is run by a commercial company. The mix servers used to build different mix cascades are operated by independent and non interrelated organizations or private individuals who all publish their identity. The service is available for free with several limitations, like the maximum download speed. In addition, there are different premium rates without these limitations that differ with regard to duration and included data volume. Thus, JonDonym offers several different tariffs and is not based on donations. The actual number of users is not predictable since the service does not keep track of this.

Thus, we assume that users' perceptions are equal with respect to technical characteristics, but may be different with respect to trust in the services.

From a research perspective, there are some papers about JonDonym, e.g. a user study on user characteristics of privacy services [55]. However, the majority of work is about Tor. Most of the work is technical [50], e.g. on improvements such as relieved network congestion, improved router selection, enhanced scalability or reduced communication/computational cost of circuit construction [4]. Naturally, there is also lots of work about the security and anonymity properties [33, 37] and traffic correlation [34].

2.2 Research on Technology Acceptance

The field of technology adoption and use has been the subject of a multitude of previous research, yielding several competing concepts, theories, and models. Some of the most prominent models will be briefly introduced in order to create a common understanding for the following analysis and our choice for using the technology acceptance model (TAM) as the base model.

The Theory of Reasoned Action (TRA) provides the theoretical starting point of TAM. It falls back on empirical research conducted by the social psychologists Fishbein and Ajzen [22]. According to TRA, a person's behaviour is determined by that person's intention to perform this particular behaviour. The behavioural intention (BI), in turn, is influenced by his or her subjective norms (SN) and attitude toward the given behaviour (A). BI can also be viewed as a function of certain beliefs. On the one hand, attitude is related to a person's beliefs about and evaluation of the consequences of the behaviour. On the other hand, the subjective norms concerning a given behaviour are affected by normative beliefs and normative pressure. Subjective norms refer to a person's motivation to comply with persons saying whether he or she should perform the behaviour or not. Feedback loops can arise at various stages of the process, as the performance of a given behaviour can have an impact on beliefs, which in turn influences BI and hence the behaviour itself.

The Theory of Planned Behavior (TPB) by Ajzen [3] is based on the TRA. The overall structural process remains unchanged, i.e. BI is influenced by several components and in turn influences the performance of a behaviour. Nevertheless, it was created as an extension of the TRA integrating the addition of perceived behavioural control (PBC). In practical terms, this denotation refers to a person's perception regarding the ease or difficulty of performing a given behaviour in a given

situation. Consequently, PBC is assumed to depend on the extent to which required resources and opportunities are available. PBC can have an impact on behaviour in two ways. First, indirectly through its influence on BI and its relationship with A and SN. Secondly, together with BI, PBC can be used directly for predicting behavioural achievement.

Based on the TRA and TPB, TAM was developed in 1985 by Davis [18]. The model specifically focuses on the user acceptance of information systems. Similar to TRA, TAM hypothesizes that system use is determined by BI to use. However, it differs from the former model, as BI is jointly influenced by a person's overall attitude towards the use of the technology (A) and the perceived usefulness (PU). Subjective perceptions regarding the system's ease of use are theorized to be fundamental determinants of the system use, too. They directly influence A and PU. Again, PU refers to the extent to which a system would enhance a person's job performance within an organizational context. Perceived ease of use (PEOU) is the degree of effort needed to use the system. Furthermore, external variables affect one's attitude and behaviour indirectly through their impact on PU and PEOU [20]. TAM has been the subject of various studies and extensions whereas PETs were, to the best of our knowledge, seldom considered as a research object in the context of TAM (e.g. the paper by Benenson et al. [8] is based on TAM for the case of anonymous credentials). However, the model is well suited for our case of explaining the behavioral intention and actual use behavior of PETs due to the following reasons. First, the model and the respective constructs are widely tested in the literature and the base model provides valid and reliable measures of the above mentioned variables. Thus, we argue that these constructs provide an appropriate basis for explaining technology acceptance of PETs. Second, the model is parsimonious, i.e. there are relatively few constructs necessary to explain a relatively large share of the variance in the target constructs. This makes it possible to add technology-specific variables (in our case for PETs) without overspecifying the model and minimizing an overspecification bias. We adapt the original constructs of TAM to the case of PETs by specifying perceived usefulness as the usefulness of a PET to protect the user's privacy. We argue that this definition is reasonable for our examplary PETs (Tor and JonDonym) since they enable users to do multiple tasks while privacy protection is the evident goal when using them. This perception regarding the usefulness to protect the user's privacy is therefore theorized to be crucial when deciding to use a PET. In summary, we argue that our adapted TAM model serves as an appropriate theoretical underlying for answering our research questions and contribute to our understanding regarding the main factors influencing individuals' use behavior of PETs.

2.3 Related Work

Previous non-technical work on PETs mainly considers usability studies and does not primarily focus on technology acceptance of these technologies. For example, Lee et al. [39] assess the usability of the Tor Launcher and propose recommendations to overcome the found usability issues. In a qualitative study, Forte et al. [24] examine perceived risks and privacy concerns of Tor users and Wikipedia editors who are concerned about their privacy. Previous related work investigates privacy concerns and trust with respect to JonDonym [30] and Tor [31] based on Internet users' information privacy concerns (IUIPC) [40]. Comparable studies to the study at hand with respect to the underlying theory of technology acceptance are the ones by Benenson et al. [8, 9] and Krontiris et al. [38] who investigate acceptance factors for an anonymous credential service. However, in their case the anonymous credential service is integrated into a course evaluation system. Thus, the users of their anonymous credential service had a clearly defined primary task (evaluation of the course system) and a secondary task (ensure privacy protection). Benenson et al. focused on the measurement of the perceived usefulness of the anonymous credential system (the secondary goal), but state that considering the perceived usefulness for the primary goals as well, may change the relationship between the variables in their model [8]. In contrast to their study, we examine a standalone PET, and thus can focus on privacy protection as the primary goal of the users with respect to the PET. Compared to the previous studies, Brecht et al. [11] focus on no specific anonymization service in their analysis on acceptance factors. In addition, they do not base their model on classical technology acceptance variables like we do in this paper.

3 Methodology

In the following subsections, we discuss the research model and hypotheses based on the extended TAM, the questionnaire and the data collection process. In addition, we provide a brief overview of the employed quantitative statistical analysis approach.

3.1 Research Model and Hypotheses

PETs are structurally different compared to technologies used in the job context or pleasure-oriented (hedonic) information systems like games. Therefore, the research hypotheses and the model must be derived according to the properties of the specific technology (see Table 3 for the differences of the results between Tor and JonDonym [29]).

In general, it is obvious to users what a certain technology does. For example, if users employ a spreadsheet program in their job environment, they will see the immediate result of their action when the program provides them a calculation. The same holds for pleasure-oriented technologies which provide an immediate feedback to the user during the interaction. However, this interaction and feedback structure is different with PETs. Anonymity is the main goal which a user can achieve by using PETs. However, most PETs are designed to not harm the user experience. Besides some negative side effects such as a loss of speed during browsing the internet or an increasing occurrence of CAPTCHAs [15], the user may not be able to detect the running of the PET at all (which would be the optimal characteristic of a PET). The direct effects of the increased anonymity in general go undetected since they consist of long term consequences, e.g. different advertisements, unless the user visits special websites with anonymity tests or showing the internet address of the request. In summary, the main impact of a PET is not immediately tangible for the user.

Therefore, perceptions about the achieved impact of using the technology should be specifically incorporated in any model dealing with drivers of *use behavior*. This matches the observation that most users do not base their decisions on any kind of formal (technical or mathematical) anonymity measurement. Thus, we adapted a formerly tested and validated construct named "perceived anonymity" to the case of the PETs Tor and JonDonym [8]. The construct mainly asks for the perceptions of users about their level of anonymity achieved by the use of the PET. Due to the natural importance of anonymity for a PET, we argue that these perceptions will have an important effect on the *trust in the technology*. Thus, the more users think that the PET will create anonymity during their online activities, the more they will trust the PET (H1a). Creating anonymity for its users is the main purpose of a PET. Thus, we hypothesize that the *perceived anonymity* has a positive effect on the *perceived usefulness of the PET to protect the users' privacy* (H1b).

H1a: Perceived anonymity when using PETs has a positive effect on trust in PETs.

H1b: Perceived anonymity when using PETs has a positive effect on the perceived usefulness of PETs to protect the users' privacy.

Trust is a diverse concept integrated in several models in the Information Systems (IS) domain. It is shown that different trust relationships exist in the context of technology adoption of information systems [54]. Trust can refer to the technology (in our case PETs (Tor and JonDonym)) as well as to the service provider. Since the non-profit organization of Tor evolved around the service [56], it is rather difficult for users to distinguish which label refers to the technology itself and which refers to the organization. The same holds for JonDonym since JonDonym is the only main service offered by the commercial company JonDos. Therefore, we argue that it is rather difficult for users to distinguish which label refers to the technology itself and which refers to the company. Thus, we decided to ask for *trust in the PET* (Tor and JonDonym, respectively), assuming that the difference to ask for trust in the organization / company is negligible.

Literature shows that trust in services enables positive attitudes towards interacting with these services [44]. Applying this logic to the case of technologies, we hypothesize that a higher level of trust in a given technology causes a stronger *behavioral intention to use* this technology (H2a). Besides this direct effect on use intentions, trust influences the perceived usefulness of a given technology. Thus, we argue that the higher the trust in the PET, the higher is the level of *perceived usefulness of protecting the user's privacy* (H2b). Lastly, we hypothesize that *trust in PETs* has a positive effect on the *perceived ease of use of PETs* (H2c). Previous literature supports this hypothesis, indicating that a higher level of trust in a given technology decreases the need to understand each and every detail of the technology [14]. This is especially relevant for the case of PETs since they represent a kind of technology with a relatively high level of complexity (e.g. compared to pleasure-oriented information systems).

H2a: Trust in PETs has a positive effect on the behavioral intention to use the technology.

H2b: Trust in PETs has a positive effect on the perceived usefulness of protecting the user's privacy.

H2c: Trust in PETs has a positive effect on the perceived ease of use of PETs.

The theoretical underlying of hypotheses H3, H4a, H4b and H5 is adapted from the original work on TAM by Davis [18, 19] since PETs are not different to other technologies with regard to the relationships of *perceived usefulness*, *perceived ease*, *behavioral intention to use* and *actual use behavior*. However, *perceived usefulness* refers explicitly to privacy protection as it is the sole purpose of the technology. The rationale for hypotheses 3 and 4a are straightforward. The higher the *perceived usefulness* and *ease of use* of a given technology, the stronger the *behavioral intention to use* this technology. Literature indicates that *perceived ease of use* itself has a positive effect on the *perceived usefulness* of a technology (H4b). Improvements in *ease of use* contribute to efficiency gains and enable users of a given technology to accomplish the same goals with less effort [18, 19]. We argue that this rationale also holds for PETs, since a PET which is easy to use requires less mental effort to fulfill the goal of protecting user's privacy. Research on the relationship between *behavioral intention* and *actual use behavior* consistently indicates

that there is a positive relationship between the two variables, where *behavioral intention* has a positive effect on *actual use behavior* [22, 52]. We assume that this relationship is also apparent for the case of PETs (H5). In summary, we hypothesize:

H3: The perceived usefulness of protecting the user's privacy has a positive effect on the behavioral intention to use the technology.

H4a: Perceived ease of use has a positive effect on the behavioral intention to use the technology.

H4b: Perceived ease of use has a positive effect on the perceived usefulness of protecting the user's privacy.

H5: The behavioral intention to use PETs has a positive effect on the actual use behavior.

These hypotheses constitute the research model illustrated in Figure 1.

3.2 Questionnaire and Data Collection

The questionnaire constructs are adapted from different sources. *Perceived ease of use* (PEOU) and *perceived usefulness* are adapted from Venkatesh and Davis [59], *behavioral intention* (BI) is adapted from Venkatesh et al. [60], *trust in the PET service* is adapted from Pavlou [44] and *perceived anonymity* is adapted from Benenson et al. [8]. The former constructs are measured based on a seven-point Likert scale, ranging from "strongly disagree" to "strongly agree". The *actual use behavior* is measured with a ten-item frequency scale [49]. The adapted questionnaire items can be found in Table 1. These items are solely used for the quantitative analysis in Section 5. Besides these questions, we asked participants for their age, education and gender. However, we cannot present a reliable overview of these variables since they were not mandatory to fill out. This was done on purpose since we assumed that most of the participants are highly sensitive with respect to their personal data and could potentially react to mandatory demographic questions by terminating the survey. Consequently, the demographics are incomplete to a large extent. Therefore, we had to resign from a discussion of the demographics in our research context.

We conducted the studies with German and English-speaking users of Tor and JonDonym. For each service, we administered two questionnaires. All items for the German questionnaire had to be translated into German since all of the constructs are adapted from English literature. To ensure content validity of the translation, we followed a rigorous translation process: We translated the English questionnaire into German with the help of a certified translator (translators are standardized by the DIN EN 15038 norm). The German version was then given to a second independent certified translator who retranslated the questionnaire to English. This step was done to ensure the equivalence of the translation. Last, a group of five academic colleagues checked the equivalence of the two English versions. All items were found to be equivalent.

Since we investigate the drivers of the *use behavior of PETs*, we collected data from actual users of the PETs. We installed the surveys on a university server and managed it with the LimeSurvey [51]. For Tor, we distributed the links to the English and German version over multiple channels on the internet. Although there are 2,000,000 to 8,000,000 active users of the service, it was relatively difficult to gather the necessary number of complete answers for a quantitative analysis. Thus, to foster future research about Tor users, we provide an overview of every distribution channel in the appendix. In sum, 314 participants started the questionnaire (245 for the English version, 40 for the English version posted in hidden service forums and 29 for the German version). Of those 314 approached participants, 135 (105 for the English version, 13 for the English version posted in hidden service forums and 17 for the German version) filled out the questionnaires completely. After deleting all participants who answered a test question in the middle of the survey incorrectly, 124 usable data sets remained for the following analysis. The test question simply asked participants to select a specified answer in a given set. Questions like this are usually added to questionnaires to check for the awareness of the participants and avoid participants just clicking through the survey without carefully reading the questions.

For JonDonym, we distributed the links to the English and German version with the beta version of the JonDonym browser and published them on the official JonDonym homepage. This made it possible to address the actual users of the PET in the most efficient manner. 416 participants started the questionnaire (173 for the English version and 243 for the German version). Of those 416 approached participants, 141 (53 for the English version and 88 for the German version) remained after deleting unfinished sets and all participants who answered a test question in the middle of the survey incorrectly. In total, our sample consists of 265 complete answers.

We also addressed potential ethical issues of the user survey. The ethics board of the authors' university provides an extensive checklist which qualifies our study as exempt for an ethics review. However, in order to inform participants about our data collection process we provided information about the related research project and the goal of the study (improve PETs and investigate their acceptance factors). Furthermore, we stated that all answers are anonymous (e.g. no saving of IP addresses), that all answers are stored on a German server and that by participating in the survey, participants agree that their answers are used for scientific publications, research publications and a PhD thesis. We provided an open-text-field for

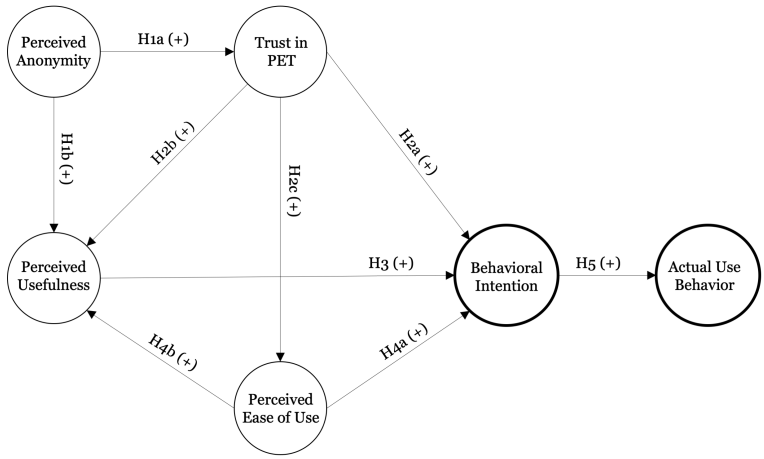


Fig. 1. Research model showing the structural model with the research hypotheses

feedback and a researcher's e-mail address for further questions and requests at the end of the survey.

3.3 Statistical Analysis Approach

We hypothesize that perceived anonymity and trust in the PET, along with the standard variables drawn from the TAM (cf. Section 2.2), are measurable underlying constructs that influence the adoption of Tor and JonDonym. To test this, we use the questionnaire described in Section 3.2 to measure these constructs, and apply a standard statistical analysis approach called structural equation modelling (SEM) to assess our research model and the corresponding hypotheses regarding the cause-effect relationships among these constructs. SEM can reveal how much of the variance in the dependent variables (effects) can be explained by the independent variables (causes). There are two main approaches for SEM, namely covariance-based SEM (CB-SEM) and partial least squares SEM (PLS-SEM). Since our research goal is to predict the dependent variables (effects) *behavioral intention* and *actual use behavior* of PETs and maximize the explained variance for these dependent variables, we use PLS-SEM [27] for our analysis (Hair et al. extensively discuss on the use of PLS-SEM [26]).

4 Validity and Reliability Testing

We tested our model (cf. Section 3) using SmartPLS version 3.2.7 [48]. Before looking at the result of the structural model and discussing its implications, we discuss the measurement model, and check for the reliability and validity of our results. This is a precondition of being able to interpret the results of the structural model. Furthermore, it is recommended to report the computational settings. For the PLS algorithm, we chose the suggested path weighting scheme with a maximum of 300 iterations and a stop criterion of 10^{-7}. For the bootstrapping procedure, we used 5000 bootstrap subsamples and no sign changes as the method for handling sign changes during the iterations of the bootstrapping procedure [26]. We met the suggested minimum sample size with 265 datasets considering the threshold of ten times the number of structural paths headed towards a latent construct in the model [27].

4.1 Measurement Model Assessment

As the model is measured solely reflectively, we need to evaluate the internal consistency reliability, convergent validity and discriminant validity to assess the measurement model properly [27]. Internal consistency reliability (ICR) measurements indicate how well certain indicators of a construct measure the same latent phenomenon. Two standard approaches for assessing ICR are Cronbach's α and the composite reliability. The values of both measures should be between 0.7 and 0.95 for research that builds upon accepted models. Values of Cronbach's α are seen as a lower bound and values of the composite reliability as an upper bound of the assessment [26]. Table 1 includes the ICR of the variables in the last two rows. It can be seen that all values for Cronbach's α and the composite

reliability are above the lower threshold of 0.7 and no value is above 0.95. In sum, ICR is established for our variables.

In a next step, we assess the convergent validity to determine the degree to which indicators of a certain reflective construct are explained by that construct. For that, we calculate the outer loadings of the indicators of the constructs (indicator reliability) and evaluate the average variance extracted (AVE) [26]. Loadings above 0.7 imply that the indicators have much in common, which is desirable for reflective measurement models. Table 1 shows the outer loadings with grey background on the diagonal. All loadings are higher than 0.7. Convergent validity for the construct is assessed by the AVE. AVE is equal to the sum of the squared loadings divided by the number of indicators. A threshold of 0.5 is acceptable, indicating that the construct explains at least half of the indicators' variance. The first column of Table 2 presents the constructs' AVE. All values are above 0.5, demonstrating convergent validity.

The next step for assessing the measurement model is the evaluation of discriminant validity. It measures the degree of uniqueness of a construct compared to other constructs. Two approaches are used for investigating discriminant validity. The first approach, assessing cross-loadings, is dealing with single indicators. All outer loadings of a certain construct should be larger than its cross-loadings with other constructs [26]. Table 1 illustrates the cross-loadings as off-diagonal elements. All cross-loadings are smaller than the outer loadings, fulfilling the first assessment approach of discriminant validity. In the second approach, we compare the square root of the constructs' AVE with the correlations with other constructs. The square root of the AVE of a single construct should be larger than the correlation with other constructs (Fornell-Larcker criterion). Table 2 contains the square root of the AVE as on-diagonal values. All values fulfill the Fornell-Larcker criterion, indicating discriminant validity.

The last step of the measurement model assessment is to check for common method bias (CMB). CMB can occur if data is gathered with a self-reported survey at one point in time in one questionnaire [41]. Since this is the case in our research design, we test for CMB. An unrotated principal component factor analysis is performed with the software package STATA 14.0 to conduct the Harman's single-factor test to address the issue of CMB [47]. The assumptions of the test are that CMB is not an issue if there is no single factor that results from the factor analysis or that the first factor does not account for the majority of the total variance. The test shows that four factors have eigenvalues larger than 1 which account for 72.04% of the total variance. The first factor explains 46.51% of the total variance. Thus, no single factor emerged and the first factor does not explain the majority of the variance. Hence, we argue that CMB is not likely to be an issue.

4.2 Structural Model Assessment

We first test for possible collinearity problems before discussing the results of the structural model. Collinearity is present if two predictor variables are highly correlated with each other. This is important since collinearity can otherwise bias the results heavily. To address this issue, we assess the inner variance inflation factor (inner VIF). All VIF values above 5 indicate that collinearity between constructs is present [26]. For our model, the highest VIF is 1.892. Thus, collinearity is apparently not an issue.

We also assessed the predictive relevance of the two added variables for *behavioral intention* and *actual use behavior* in order to assess whether they are important enough to be included in the model. A simple measure for the relevance of *perceived anonymity* and *trust* is to delete both variables and run the model again. The results show that the R^2-value for *behavioral intention* decreases to 41.9% (= 5.8 percentage points less). Thus, without the two new variables the explained variance for *behavioral intention* decreases by 12.2%. A more advanced measure for predictive relevance is the Q^2 measure. It indicates the out-of-sample predictive relevance of the structural model with regard to the endogenous latent variables based on a blindfolding procedure [26]. We used an omission distance d=7. Recommended values for d are between five and ten. Furthermore, we report the Q^2 values of the cross-validated redundancy approach, since this approach is based on both the results of the measurement model as well as of the structural model. Detailed information about the calculation is given by Chin [13]. For our model, Q^2 is calculated for *behavioral intention* and *use behavior*. Values above 0 indicate that the model has the property of predictive relevance. Omitting both new variables leads to a decrease of Q^2 for *behavioral intention* from 0.336 to 0.293. R^2 as well as Q^2 did not change for *actual use* when deleting the new variables, since there is no direct relation from these constructs to *actual use*.

5 Quantitative Analysis Results

We present the results of our quantitative analysis in this section. First, we discuss the path estimates and the R^2-values for our extended technology acceptance model. Second, we conduct a multigroup analysis in order to investigate potential differences in the path estimates between Tor and JonDonym.

Constructs	BI	PEOU	PA	$Trust_{PETs}$	PU	USE
BI1. I intend to continue using the PET[1] in the future.	0.884	0.499	0.537	0.573	0.602	0.322
BI2. I will always try to use the PET[1] in my daily life.	0.830	0.409	0.350	0.408	0.372	0.319
BI3. I plan to continue to use the PET[1] frequently.	0.931	0.487	0.439	0.545	0.534	0.408
PEOU1. My interaction with the PET[1] is clear and understandable.	0.503	0.825	0.281	0.386	0.410	0.153
PEOU2. Interacting with the PET[1] does not require a lot of my mental effort.	0.390	0.826	0.232	0.259	0.361	0.178
PEOU3. I find the PET[1] to be easy to use.	0.450	0.911	0.233	0.316	0.386	0.211
PEOU4. I find it easy to get the PET[1] to do what I want it to do.	0.468	0.882	0.338	0.382	0.473	0.232
PA1. The PET[1] is able to protect my anonymity in during my online activities.	0.488	0.311	0.899	0.593	0.641	0.103
PA2. With the PET[1] I obtain a sense of anonymity in my online activities.	0.437	0.259	0.885	0.609	0.616	0.143
PA3. The PET[1] can prevent threats to my anonymity when being online.	0.418	0.276	0.871	0.544	0.582	0.126
$Trust_{PETs}1$. The PET[1] is trustworthy.	0.513	0.348	0.642	0.891	0.608	0.115
$Trust_{PETs}2$. The PET[1] keeps promises and commitments.	0.557	0.386	0.581	0.921	0.568	0.139
$Trust_{PETs}3$. I trust the PET[1] because they keep my best interests in mind.	0.509	0.335	0.556	0.895	0.545	0.166
PU1. Using the PET[1] improves the performance of my privacy protection.	0.349	0.338	0.459	0.442	0.782	0.130
PU2. Using the PET[1] increases my level of privacy.	0.559	0.433	0.668	0.626	0.934	0.210
PU3. Using the PET[1] enhances the effectiveness of my privacy.	0.439	0.429	0.604	0.499	0.882	0.136
PU4. I find the PET[1] to be useful in protecting my privacy.	0.628	0.456	0.662	0.627	0.896	0.225
USE. Please choose your use frequency[2] of the PET[1].	0.398	0.225	0.140	0.155	0.206	1.000
Cronbach's α	0.859	0.885	0.862	0.886	0.898	-
Composite Reliability	0.914	0.920	0.916	0.929	0.929	-

BI: Behavioral Intention PEOU: Perceived Ease of Use PA: Perceived Anonymity USE: Actual Use Frequency

PU: Perceived Usefulness of Protecting Users' Privacy [1]Tor/JonDonym [2]10-point scale from "Never" to "All the time"

Table 1. Loadings and cross-loadings of the reflective items and ICR measures

Constructs (AVE)	BI	PA	PEOU	PU	$Trust_{PETs}$
BI (0.780)	0.883				
PA (0.783)	0.507	0.885			
PEOU (0.743)	0.530	0.319	0.862		
PU (0.766)	0.579	0.693	0.477	0.875	
Trust (0.814)	0.583	0.658	0.396	0.636	0.902
USE	0.398	0.140	0.225	0.206	0.155

Table 2. Discriminant validity and construct correlations

5.1 Technology Acceptance Factors of PETs

Figure 2 presents the results of the path estimations and the R^2-values of the target variables *behavioral intention* and *actual use behavior*. In addition, we provide the R^2-values for *trust*, *perceived ease of use* and *perceived usefulness*. R^2-values are weak with values around 0.25, moderate with 0.50 and substantial with 0.75 [27]. Based on this classification, the R^2-value for *behavioral intention* is moderate in size and weak for the variable *actual use behavior*. Our model explains 47.7% of the variance in the *behavioral intention to use the PET* and 15.8% of the variance of the *actual use behavior*.

In the Tor survey, several participants answered that they never use Tor (21 participants answered "never" to the question about their use frequency of Tor). This statement of these 21 participants is in contrast to their answer to a question in which we asked participants how many years they are using Tor. Here, the respective participants stated that they used Tor for six years (median of 6 years and an average of 6.87 years). The correlation coefficient between the years of using Tor and the use frequency is very small and negative with -0.0222. These 21 answers massively bias the results for the relationship between *behavioral intention* and *actual use behavior* (the median value of use frequency is 5). However, we cannot explain why the participants answered like this. They either misunderstood the question, answered it intentionally like this to disguise their activity with Tor or found the scale for use behavior inappropriate. This might be due to the fact that the scale only contains "once a month" as the lowest use frequency besides "never". It might be possible that these 21 users use Tor only a few times per year or that they used Tor some years ago and have not used it again since then. Therefore, they might have chosen never as an answer. However, we used an established scale to measure use behavior [49], but recommend to consider this issue in future research with a similar context. For JonDonym, we did not observe this issue. The respective path coefficients are shown in Table 3. The effect size between *behavioral intention* and *actual use* is 0.679 for JonDonym and 0.179 for Tor.

Three main drivers of perceived usefulness of PETs

The explained variance of *perceived usefulness* is 58.4%, indicating that the three variables, *perceived anonymity*, *trust* and *perceived ease of use* explain almost two-thirds of the variance of this construct. Thus, we identified three major drivers of users' perceptions with regard to the usefulness of a privacy-enhancing technology. This result shows that the two

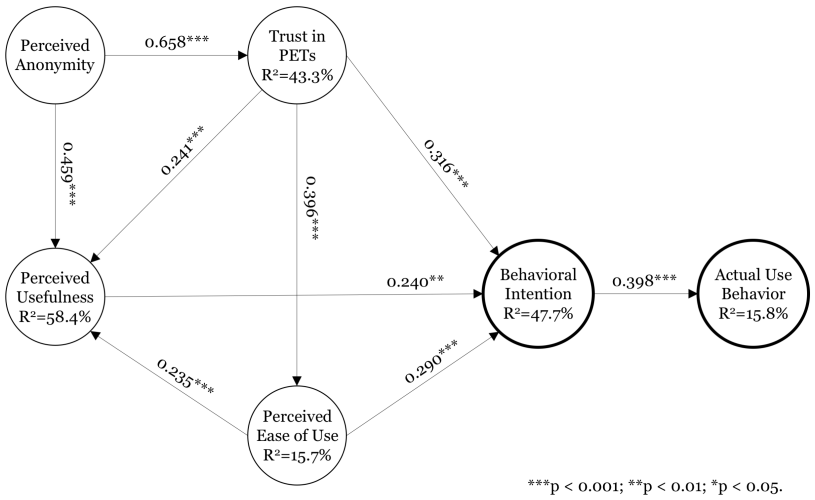


Fig. 2. Research model with path estimates and R^2 values of the structural model for PETs

	Relationships	Path coeff. original (JonDonym)	Path coeff. original (Tor)	P-values (JonDonym)	P-values (Tor)	Diff. path coeff. (| JonDonym - Tor |)	P-values (JonDonym vs Tor)
H1a	PA → Trust_{PETs}	0.597	0.709	< 0.001	< 0.001	0.112	0.865
H1b	PA → PU	0.543	0.369	< 0.001	< 0.001	0.174	0.088
H2a	Trust_{PETs} → BI	0.416	0.232	< 0.001	0.010	0.184	0.064
H2b	Trust_{PETs} → PU	0.173	0.304	0.035	0.008	0.131	0.823
H2c	Trust_{PETs} → PEOU	0.378	0.431	< 0.001	< 0.001	0.053	0.657
H3	PU → BI	0.183	0.300	0.046	0.002	0.117	0.805
H4a	PEOU → BI	0.206	0.371	0.011	< 0.001	0.165	0.929
H4b	PEOU → PU	0.182	0.300	0.039	< 0.001	0.118	0.830
H5	BI → USE	0.679	0.179	< 0.001	0.029	0.500	< 0.001

BI: Behavioral Intention PEOU: Perceived Ease of Use PA: Perceived Anonymity USE: Actual Use Frequency
PU: Perceived Usefulness of Protecting Users' Privacy

Table 3. Results of the MGA-analysis (grey background indicates statistical significance at least at the 10% level)

newly added variables are important antecedents in the technology acceptance model which should be considered in future work on this topic. The strongest effect is exerted by the users' *perceived anonymity* provided by the service (H1b confirmed). This result is not surprising considering that providing anonymity is the main goal of a PET. In addition, *perceived anonymity* has a relatively strong and statistically significant effect on *trust* (H1a confirmed). Thus, users' *trust in PETs* is mainly driven by their perceptions that the service can create anonymity (R^2-value of Trust_{PETs} equals 43.3%).

Trust in PETs is the most important factor

As hypothesized in H2a - H2c, *trust* has a significant positive effect on the *behavioral intention to use the PET*, the *perceived usefulness* and the *perceived ease of use*. Therefore, *trust* emerges as a highly relevant concept when determining the drivers of users' *use behavior of PETs*. Among the factors influencing *behavioral intention*, it has the strongest effect size (0.316). As discussed earlier, hypotheses H3 - H5 are adapted from the original work on TAM [18, 19] and can be confirmed for the case of PETs.

Significant total effects of trust and perceived anonymity

Since the effects of *perceived anonymity* and *trust* on *behavioral intention* and the *actual use behavior* are partially indirect, we determine and analyze the total effects for these variables (cf. Table 4). It can be seen that the total effects for *behavioral intention* are relatively large and highly statistically significant. Thus, *perceived anonymity* and *trust* strongly influence the target variable *behavioral intention*. Due to the

Total effect	Effect size	P-value
PA → BI	0.446	< 0.001
PA → USE	0.177	< 0.001
Trust_{PETs} → BI	0.511	< 0.001
Trust_{PETs} → USE	0.203	< 0.001

Table 4. Total effects for perceived anonymity and trust in PETs

discussed bias in the construct USE, the total effects for this variable are comparably small.

5.2 Multigroup Analysis

After the analysis of the whole data sample, we split the data set into two parts and analyze the results for Tor and JonDonym separately. For that, we conduct a multigroup analysis and test whether there are statistically significant differences for each of the hypotheses.

Since JonDonym and Tor are different with respect to the pricing schemes and the organizational structure of the providers, we are interested whether there are significant differences in the hypothesized relationships between the variables. For that purpose, we conducted a multigroup analysis in SmartPLS (cf. Table 3). We use a less conservative level of statistical significance of 10% in this table since the p-value is sensitive to the relatively small sample sizes when comparing results for Tor and JonDonym. Thus, we provide this level of statistical significance in this analysis to indicate potential statistically significant differences between the effects for Tor and JonDonym. In addition, the oftentimes referenced statistical significance level of 5% only indicates a "convenient" threshold for judging statistical significance [23] and can be considered a rule of thumb.

Trust is less important for Tor than for JonDonym

The results indicate that all relationships are similar for both PETs with respect to direction of the effect and effect size (see the path coefficients for both PETs). This supports the assumption that Tor is comparable to JonDonym from a user's perspective. Only three relationships are significantly different for the two technologies (p-value of difference smaller than 0.1). First, the effect of *perceived anonymity* on *perceived usefulness* is weaker for Tor than for JonDonym. Furthermore, *trust in the PET* is significantly less important for Tor than for JonDonym.

Differences in these relationships can have many causes. Among others, Tor exists longer and has significantly more users. However, the results are especially interesting when considering the structures of the two organizations. Tor has a more community-oriented structure based on donations, whereas JonDonym is operated by a profit-oriented company which charges money for the unlimited use of the PET [35]. Thus, users possibly focus more on the trust in the PET if it is operated by a commercial company, which leads to a stronger influence of trust on the use intentions and behaviors.

In contrast to this, Tor might be perceived as a technology that is based on the community which operates the used servers voluntarily without financial intentions. This leads to a wide distribution of the infrastructure and trust in the service is not needed from a technical point of view since the communication can only be intercepted if each server is controlled by one attacker. Therefore, users might perceive that the need for trust is not as important as if a profit-oriented company operates the PET.

6 Qualitative Analysis Results

We augment our quantitative results from the previous section with a qualitative analysis of answers to five open questions included in the questionnaires. By that, we provide deeper insights into certain aspects of the quantitative analysis from Section 5 and hints to relevant questions for future work. We show the questions and the number of answers to them in Table 5. These numbers exclude answers as "I don't know", "no" and so on. Two researchers analyzed the statements independently from each other and abstracted the invidiual answers to codes. Codes summarize the data and present different dimensions of a concept. For example, we find that *usability* is an important concept for both technologies. However, the results indicate that *usability* can be both a negative as well as a positive characteristic, depending on the user and the respective context. For example, the code "usability" joins negative as well as positive perceptions of users.

We do the coding of the 626 statements to the open questions in two stages. We use a coding method from sociology [12, 25], which comprises two or three coding phases, namely initial coding, axial coding and focused coding. We only use initial and focused coding since this level of structuring is sufficient for our data [12]. First, we initially code each of the statements. These initial codes in itself provide a sorting and structuring for the data. Initial codes represent topics that occur frequently in the data, i.e. topics often mentioned by participants. In our case, we decide to name these codes "Subconcepts" in our results since they already provide one level of abstraction. After the initial coding phase, we compare the different codings of the researchers and discussed the individual codes. Thereby, we agreed upon certain subconcepts which were similar or the same but expressed differently by the coders. In a next step, we calculated the intercoder reliability. We did not use a common codebook or a predefined set of codes to do the initial coding. Therefore, known reli-

Questions	Number of answers for	
	JonDonym	Tor
1. Do you have any concerns about using JonDonym / Tor?	56	85
2. Under which circumstances would you choose one of the premium tariffs? (JonDonym)	76	not applicable
3. Which additional features would you like to have at your current tariff? (JonDonym)	32	
3. Which additional features would you like to have for Tor?		124
4. Why would you recommend JonDonym / Tor?	122	102
5. Why would you not recommend JonDonym / Tor?	11	18
Σ	297	329

Table 5. Open-ended questions from the survey and number of answers

ability measures as Cohen's Kappa [16] are not usable for our case since these measures are relying on predefined categories. Consequently, we use a very simple calculation in order to provide a reliability measure dividing the number of equally coded statements by the total number of statements to be coded. We had 226 matches for Tor and 242 matches for JonDonym, which yields a intercoder reliablity of 68.69% and 81,48%, respectively (cf. Table 5 for the total number of statements for each PET). Thus, the intercoder reliability is equal to 74.76% for both PETs. These numbers are relatively large considering that we coded independently from each other without agreeing to fixed subconcepts beforehand. We also count the incidents in which one of the coders had at least one more code assigned to a statement than the other coder in order to provide more transparency of our coding process. This happened 52 times (coder 1 had 29 times more codes, coder 2 had 23 times more codes) for Tor and 44 times for JonDonym (coder 1 had 27 times more codes, coder 2 had 17 times more codes). These instances are counted towards the mismatches in the intercoder reliability measures.

In the second step, we structured the most occuring themes in these initial codes and came up with the focused codes. We name these codes "Concepts" in Table 6 since we find that users primarily make statements about either technical issues, about their beliefs and perceptions or about economic issues.

During the coding, we saw that there are certain subconcepts that hold for both, Tor and JonDonym. However, there are also subconcepts which are different for both PETs or nonexistent in the data for either one of the technologies. Therefore, we illustrate these differences separately in columns four and five of Table 6. We provide quotes from the statements for each concept, except for "Costs" and "Payment methods" since they are rather straightforward and users just stated that JonDonym should be cheaper and offer certain payment methods mentioned in the table.

Similar subconcepts to quantitative model

The results include four subconcepts which can be found in the investigated model of the quantitative part (Section 5). Participants mention *usability*, *performance*, *anonymity* and *trust* oftentimes in the context of concerns or why they would or would not recommend the respective PET. As mentioned before, these concepts are not tied to a certain positive or negative interpretation. This becomes obvious when looking at the exemplary quotes in the table.

Usability positively influences use behavior

Usability is mentioned most of the times in the context of a positive factor influencing the use. This means, if a PET is easy to use, users will prefer to use it (**Tor.5**, **Jon.5**). In contrast, participants mentioned for both PETs that they would like to have a better documentation in order to enhance the usability (**Tor.4**, **Jon.4**). We also find another interesting dimension for *usability* in the data. Some participants stated that missing knowledge about the correct use of the PET can lead to worse results with respect to privacy than without using the PET at all. This implies that some users are concerned that the degree of *ease of use* is not as high as it should be, especially considering layman users. This could lead to situations in which layman users think that the PET works properly, while it indeed does not (**Tor.6**, **Jon.6**).

Limited performance in the free version of JonDonym

The concept *performance* is only partially equivalent to *perceived usefulness* since we defined it as usefulness to protect the user's privacy. However, we argue that a PET needs to fulfill the requirement of low latency in order to be useful in the sense of protecting the users privacy. Therefore, we argue that the concept *performance* can be seen as the equivalent to the variable *perceived usefulness* in the quantitative model. It slightly differs for Tor (**Tor.7**) and JonDonym since participants only mention the issue for JonDonym when talking about the free of charge option (**Jon.7**, **Jon.8**) (the decreased performance is implemented by default for this option as a feature of the tariff [35]).

Anonymity and concerns regarding deanonymization

The concept *anonymity* is mentioned in the context of representing the main purpose of why participants use a PET (**Tor.9**, **Jon.10**). However, another dimension of this concept is a concern of being deanonymized by a variety of attackers, espe-

Concepts	Subconcepts	Common to both PETs	Specific Subconcepts for Tor	Specific Subconcepts for JD
Statements about Technical Issues	PET design	Feature Requests (**Tor.1**, **Jon.1**)	Malicious exit nodes (**Tor.2**)	Location of mix cascades (**Jon.2**)
	Compatibility	Accessibility of websites (**Tor.3**, **Jon.3**)		
	Usability	Documentation (**Tor.4**, **Jon.4**) Ease of use (**Tor.5**, **Jon.5**) Missing knowledge to use it correctly (**Tor.6**,**Jon.6**)		
	Performance	Latency (**Tor.7**, **Jon.7**, **Jon.8**)		
Beliefs and Perceptions	Anonymity	Concerns about deanonymization (**Tor.8**, **Jon.9**) Reason of use (**Tor.9**, **Jon.10**)		Size of the user base (**Jon.11**)
	Consequences	Fear of investigations (**Tor.10**, **Tor.11**, **Jon.12**)	Beliefs about social effects (**Tor.13**, **Tor.14**)	
	Trust		Trust in the community (**Tor.12**)	Trust in technology (**Jon.13**)
	Substitute technologies	Best available tool (**Tor.15**, **Jon.14**)		Tor as reference technology (**Jon.3**, **Jon.8**, **Jon.11**)
Statements about Economical Issues	Costs			Lower costs, other pricing schemes (**Jon.15**)
	Payment methods			Easy, anonymous payment options (**Jon.15**)
	Use cases		Circumvent Censorship (**Tor.16**)	Willingness to pay in certain scenarios (**Jon.16**, **Jon.17**)

Tor.1 *TCP support for name resolution via Tor's DNSPort [. . .]*

Tor.2 *Many exit nodes are run by governmental intelligence organisations. Exit notes can collect unencrypted data.*

Tor.3 *It can't be used on all websites; therefore it is of limited use to me*

Tor.4 *Easy to understand instructions for users with different levels of knowledge.*

Tor.5 *Tor protects privacy while on the web and is easy to use.*

Tor.6 *An unexperienced user may not understand the technical limitations of Tor and end up losing [. . .] privacy.*

Tor.7 *Increased latency makes the experience painful at times*

Tor.8 *It may fail to provide the expected level of anonymity because of attacks which may not even be known at the time they are performed (or commonplace).*

Tor.9 *It is a key component to maintaining one's privacy when browsing on the Internet.*

Tor.10 *Tor usage "Stands out"*

Tor.11 *[. . .] having a cop boot at my door because of Tor.*

Tor.12 *An end user needs to trust the network, the persons running Tor nodes and correct implementations [. . .]*

Tor.13 *Only social backlash from people thinking that Tor is mostly used for illegal activities.*

Tor.14 *For the same reason I don't hang out in brothels, using Tor makes you look like a criminal*

Tor.15 *While not perfect, Tor is the best option for reliable low-latency anonymization*

Tor.16 *It can be used as a proxy / VPN to get past censorship*

Jon.1 *Larger number of Mix Cascades, more recent software, i.e. preconfigured browser, faster security updates*

Jon.2 *First and last server of the mix cascade should not be located in the same country*

Jon.3 *Unlike Tor, JonDonym is not blocked by some websites. (Google for example among others)*

Jon.4 *Clearer explanations and instructions for JonDoFox*

Jon.5 *Easy to use, outside the mainstream like i.e. Tor*

Jon.6 *Privacy is less than expected because of wrong configuration settings.*

Jon.7 *[. . .] Even if it is quite slow without a premium tariff*

Jon.8 *[. . .] sometimes it's a little bit to slow, but compared with Tor...*

Jon.9 *Defeat of your systems by government agencies.*

Jon.10 *It provides a minimum level of personal data protection and online safety.*

Jon.11 *Tor is better due to having a much larger user base. More users results in greater anonymity*

Jon.12 *By using the service, am I automatically marked by intelligence authorities as a potential terrorist, supporter of terrorist organizations, user [. . .] for illegal things?*

Jon.13 *How can I trust Jondonym? How can Jondonym proof that servers are trustworthy?*

Jon.14 *It appeared to be the least worst option for anonymisation when I researched anonymisation services*

Jon.15 *Fair pricing, pre-paid is an easy payment option.*

Jon.16 *For use it in a country where it's difficult surf the net*

Jon.17 *If I would use the computer for work-related tasks*

Table 6. Results of the coding for the open questions including quotes

cially government agencies (**Tor.8**, **Jon.9**) and by the fact that the anonymity set is too small because of a user base which is too small. The small user base is only mentioned as a concern by users of JonDonym (**Jon.11**).

Trust as a use factor and reason for concerns

The last concept which can be found in the quantitative model is *trust in the technology*. As for *usability*, *trust* is mentioned as a concern but also as a reason for recommending both PETs in our sample. However, the qualitative analysis reveals that the trust dimensions are slightly different between Tor and JonDoym. For Tor, participants mainly mention trust in the community (**Tor.12**), whereas the community aspect is not existent for JonDonym. For JonDonym, participants mainly focus on trust in the company and the technology (**Jon.13**). In summary, our findings related to trust support the quantitative results and strengthen our claim that *trust in the technology* is a major factor in a user's decision to use a PET. However, the results also show that future work should consider to differentiate the concept of trust and adapt it to the specific context of the PET.

New concepts emerged in the qualitative analysis

The concepts "PET design", "compatibility", "social issues", "substitute technologies" and the "statements about economical issues" are not reflected in our quantitative model. Participants still mention these concepts several times and we argue that they might be interesting to consider for future work dealing with technology acceptance of PETs.

Technical design of PETs affect concerns

"PET design" describes mainly concerns about the technical structure of the PETs which is prone to attacks (especially by government agencies). Tor and JonDonym differ in their technical structure which is reflected in the statements. Several participants mention "malicious exit nodes" as a technical issue for Tor (**Tor.2**). For JonDonym, participants are mainly concerned about the location of the mix cascades (**Jon.2**). Related to "PET design" is the concept "substitute technologies". Here, several participants state for Tor and JonDonym that the respective PET is the "best option available" amongst the existing PETs (**Tor.15**, **Jon.14**). Thus, the concern about the technical design might be compensated partially by this opinion of users. Interestingly, several other JonDonym users mention Tor several times as a comparative technology to argue about advantages of JonDonym (**Jon.3**, **Jon.8**). Participants oftentimes make this comparison in the context of deciding when they would spend money for a JonDonym premium tariff. Here, they argue that they would only do this, if Tor was not existent. This is due to costs, but also due to the larger anonymity set provided by Tor (**Jon.11**). This result implies that there are very high market entry barriers for comparable commercial PETs due to the strong market position of Tor. Related to the design of Tor and JonDonym are feature requests mentioned by participants. For example, participants ask for TCP support for Tor (**Tor.1**) and faster security updates for JonDonym (**Jon.1**).

Compatibility of PETs with websites affects adoption

"Compatibility" describes concerns and statements why participants would not recommend the PETs. They primarily mention accessibility issues with websites when using the respective PET (**Tor.3**, **Jon.3**). This is an important factor to consider for future technical improvements of the PETs and closely linked to the usability. PET developers should address this issue to foster a wider market acceptance.

Fear of investigations and adverse social effects

"Consequences" are prevalent for Tor and JonDonym users. The subconcept represents the fear of PET users that their use of PETs causes them to "stand out" (**Tor.10**) and leads to investigations by police forces or other government agencies (**Tor.11**, **Jon.12**). In addition to concerns related to governmental agencies, Tor users mentioned adverse social effects due to the use of Tor. These adverse social effects describe the belief that other members of the society think negatively about Tor. For example, participants stated that Tor is oftentimes primarily associated with illegal activities by others (**Tor.13**, **Tor.14**). This subconcept is interesting for future work dealing with the acceptance of PETs in the mass market. Layman users might be susceptible to such perceptions and therefore, avoid using a PET. Thus, marketers of PETs should stress the benefits for the user's privacy and self-determination and cleary address and explain these concerns related to possible consequences and social issues.

Importance of pricing schemes and payment methods

The last part on statements about economical issues is mainly relevant for JonDonym. The concept "costs" indicates that JonDonym users would like to have other pricing schemes which are either cheaper or include more available high-speed traffic (**Jon.15**). The concept "payment methods" is showing that PET users want a variety of (mainly anonymous) payment methods like virtual currencies or paysafecards [45] (**Jon.15**). The last concept is about "use cases" which influence the decision to use a PET at all. Censorship in certain countries is the main use scenario represented in this subconcept for Tor (**Tor.16**) and JonDonym (**Jon.16**). In addition, we find that participants would pay money for JonDonym if they were required to do sensitive, work-related tasks (**Jon.17**).

7 Discussion

We found strong effects for the influence of the *perceived anonymity* on the *behavioral intention to use the PET* (RQ1). The participants mentioned anonymity several times as the main reasons why they are using Tor or JonDonym. Therefore,

the results indicate that anonymity is one of the most important factors in the use decisions of PETs. In contrast to the findings of Benenson et al. [8], who found that *trust in the PET* has no statistically significant impact on the *intention to use the service*, we found a significant medium-sized effect of *trust in the PET* on the *behavioral intention to use* it (0.316) (RQ2). One possible explanation for the difference between the literature and our results is that the trust in the service and the trust in the service provider are perceived as equivalent in our use case, whereas in the literature trust refers solely to the technology [8]. In addition, the results of the multigroup analysis revealed that *trust in the PET* has a much stronger effect on the *use intentions* if the technology is operated by a commercial company (effect stronger for JonDonym compared to Tor) [5, 35]. However, this is only one possible explanation and there could be several other omitted variables. Still, it is an interesting starting point for future work.

Our results indicate that the *use behavior of PETs* is mainly influenced by the variables *perceived usefulness* and *perceived ease of use* as well as the newly added variables *trust* and *perceived anonymity*. This result is in line with the given statements of the participants to the open questions as well as with previous studies showing that *usability* is an important aspect for the use of this PET [11, 21].

Although we checked for several reliability and validity issues, certain limitations might impact our results. First, the sample size of 265 participants (124 for Tor and 141 for JonDonym) is relatively small for a quantitative study. However, since we reached the suggested minimum sample size for the applied method, we argue that our results are still valid. In addition, it is very difficult to gather data of actual users of PETs since it is a comparable small population that we could survey. It is also relevant to mention that we did not offer any financial rewards for the participation. Secondly, our sample is likely to be biased since our sample is by default a subset of anomymity service users who are privacy sensitive individuals relative to the rest of the population. Moreover, since they answered our survey, it could be that the respondents are the least privacy sensitive of the individuals since the most privacy sensitive individuals might not even have considered to participate in our survey. Thus, certain findings from our research might not be generalizable to a potentially larger user base. A third limitation concerns possible self-report biases (e.g. social desirability). We addressed this possible issue by gathering the data fully anonymized. Fourthly, mixing results of the German and English questionnaire could be a source of errors. On the one hand, this procedure was necessary to achieve the minimum sample size. On the other hand, we followed a very thorough translation procedure to ensure the highest level of equivalence as possible. Thus, we argue that this limitation did not affect the results to a large extent. However, we cannot rule out that there are unobserved effects on the results due to running the survey in more than one country at all. In addition, we did not control for the participants' actual or former use of different standalone PETs. This experience might have an impact on their assessments of Tor and JonDonym. Furthermore, demographic questions were not mandatory to fill out due to our assumption that these types of individuals who use Tor or JonDonym are highly cautious with respect to their privacy. Thus, we decided to go for a larger sample size considering that we might have lost participants otherwise (if demographics had to be filled out mandatorily).

Future work can build on the proposed relationships and extensions of our model to investigate the acceptance and use of other PETs in more detail. We could explain more than half of the variance in the target construct *behavioral intention* with a rather parsimonious model. For the construct *actual use behavior*, we did not find comparable high values due to the issues with the answers mentioned in Section 5. Furthermore, the analysis of the open questions shows interesting new concepts to consider in future work on technology acceptance of PETs. These concepts are about the design of the respective PET, compatibility when using it (e.g. websites not working properly), social issues, negative privacy experiences, other available solutions for privacy protecting and economic factors (only relevant for commercial applications).

In addition, it would be interesting to investigate the perceptions of non-users about PETs and compare them to actual users to figure out how the perceptions of these groups differ with respect to their influence on the *use intentions* and *actual use behavior*.

8 Conclusion

Up to now research on privacy-enhancing technologies mainly focused on the technical aspects of the technologies. In addition, to the best of our knowledge, the anonymization services Tor and JonDonym were not compared in the context of technology acceptance. However, a successful implementation and adoption of PETs requires a profound understanding of the perceptions and behaviors of actual and possible users of the technologies. Thus, with this paper we investigated actual users of existing PETs as a first step to address this research problem. Our results indicate that the basic rationale of technology use models is applicable for PETs like Tor and JonDonym as well as for other comparable privacy-enhancing technologies providing a relatively strong level of anonymization. The newly introduced variables *perceived anonymity* and *trust* improved the explanatory power of the structural model for the case of PETs and can be considered as a starting point

for comparable research problems in future work. The analysis of the open questions shows that the existing variables in our technology acceptance model can also be found as relevant concepts in the statements by the participants (*usability*, *performance*, *anonymity* and *trust*). In addition, the new concepts can be considered for future studies in this area.

Our results are a first step towards a deeper understanding of the acceptance of privacy-enhancing technologies. The results provide insights for developers and marketers to specifically address issues hindering a broader diffusion of PETs. Research in this area is a real contribution for strengthening the personal right for privacy in times of ever-increasing personal data collection in the internet.

9 Acknowledgements

This work was partially supported by German Federal Ministry of Education and Research (BMBF) [grant number 16KIS0371] and by the European Union's Horizon 2020 research and innovation program from the project CyberSec4Europe [grant agreement number 830929].

References

[1] Ruba Abu-Salma, M Angela Sasse, Joseph Bonneau, Anastasia Danilova, Alena Naiakshina, and Matthew Smith. Obstacles to the Adoption of Secure Communication Tools. In *IEEE Security & Privacy*, pages 137 – 153, 2017.

[2] Alessandro Acquisti, Roger Dingledine, and Paul Syverson. On the Economics of Anonymity Alessandro. In *International Conference on Financial Cryptography*, pages 84–102. Springer Berlin Heidelberg, 2003.

[3] Icek Ajzen. The Theory of Planned Behavior. *Organizational Behavior and Human Decision Processes*, 50(2):179–211, 1991.

[4] Mashael Alsabah and Ian Goldberg. Performance and security improvements for tor: A survey. *ACM Comput. Surv.*, 49(2):32:1–32:36, September 2016.

[5] Anonymized. Examining technology use factors of privacy-enhancing technologies: The role of perceived anonymity and trust. In *24th Americas Conference on Information Systems, AMCIS 2018, New Orleans, LA, USA, August 16-18, 2018*. Association for Information Systems, 2018.

[6] James Ball. Hacktivists in the frontline battle for the internet. https://www.theguardian.com/technology/2012/apr/20/hacktivists-battle-internet, 2012.

[7] Mathieu Bédard. The underestimated economic benefits of the internet. Regulation series, The Montreal Economic Institute, 2016. Economic Notes.

[8] Zinaida Benenson, Anna Girard, and Ioannis Krontiris. User Acceptance Factors for Anonymous Credentials: An Empirical Investigation. *14th Annual Workshop on the Economics of Information Security (WEIS)*, pages 1–33, 2015.

[9] Zinaida Benenson, Anna Girard, Ioannis Krontiris, Vassia Liagkou, Kai Rannenberg, and Yannis C. Stamatiou. User Acceptance of Privacy-ABCs: An Exploratory Study. In *Human–Computer Interaction*, pages 375–386, 2014.

[10] John J. Borking and Charles Raab. Laws, PETs and Other Technologies for Privacy Protection. *Journal of Information, Law and Technology*, 1:1–14, 2001.

[11] Franziska Brecht, Benjamin Fabian, Steffen Kunz, and Sebastian Mueller. Are You Willing to Wait Longer for Internet Privacy? In *ECIS 2011 Proceedings*, 2011.

[12] Kathy Charmaz. *Constructing Grounded Theory*. Sage Publications, London, 2nd editio edition, 2014.

[13] Wynne W. Chin. The Partial Least Squares Approach to Structural Equation Modeling. In George A. Marcoulides, editor, *Modern Methods for Business Research*, pages 295–336. Lawrence Erlbaum, Mahwah, NJ, 1998.

[14] Alina M. Chircu, Gordon B. Davis, and Robert J. Kauffman. Trust, Expertise, and E-Commerce Intermediary Adoption. In *AMCIS 2000 Proceedings*, 2000.

[15] Richard Chirgwin. CloudFlare shows Tor users the way out of CAPTCHA hell, 2016.

[16] Jacob Cohen. Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit., 1968.

[17] L. F. Cranor and S. Garfinkel. *Security and Usability: Designing Secure Systems that People Can Use*. O'Reilly, Farnham, 2008.

[18] F.D. Davis. A Technology Acceptance Model for Empirically Testing New End-User Information Systems: Theory and Results. *Massachusetts Institute of Technology*, 1985.

[19] F.D. Davis. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly*, 13(3):319–340, 1989.

[20] F.D. Davis, Richard P. Bagozzi, and Paul R. Warshaw. User Acceptance of Computer Technology: a Comparison of Two Theoretical Models. *Management Science*, 35(8):982–1003, 1989.

[21] Benjamin Fabian, Florian Goertz, Steffen Kunz, Sebastian Müller, and Mathias Nitzsche. Privately Waiting – A Usability Analysis of the Tor Anonymity Network. In *AMCIS 2010 Proceedings*, 2010.

[22] Martin Fishbein and Icek Ajzen. *Belief, Attitude, Intention and Behavior: An Introduction to Theory and Research*. Addison-Wesley, Reading, MA, 1975.

[23] R. A. Fisher. *Statistical Methods for Research Workers*. Oliver & Boyd, Edinburgh, 14 edition, 1970.

[24] Andrea Forte, Nazanin Andalibi, and Rachel Greenstadt. Privacy, anonymity, and perceived risk in open collaboration: A study of tor users and wikipedians. In *CSCW*, pages 1800–1811, 2017.

[25] Barney G. Glaser and Anselm L. Strauss. *The Discovery of Grounded Theory*. Aldine Pub., Chicago, 1967.

[26] J. Hair, G. Tomas M. Hult, Christian M. Ringle, and Marko Sarstedt. *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)*. SAGE Publications, 2017.

[27] J. Hair, Christian M. Ringle, and Marko Sarstedt. PLS-SEM: Indeed a Silver Bullet. *Journal of Marketing Theory and Practice*, 19(2):139–152, 2011.

[28] David Harborth, Dominik Herrmann, Stefan Köpsell, Sebastian Pape, Christian Roth, Hannes Federrath, Dogan Kes-

dogan, and Kai Rannenberg. Integrating privacy-enhancing technologies into the internet infrastructure. *arXiv preprint arXiv:1711.07220*, 2017.

[29] David Harborth and Sebastian Pape. Examining Technology Use Factors of Privacy-Enhancing Technologies: The Role of Perceived Anonymity and Trust. In *Twenty-fourth Americas Conference on Information Systems (AMCIS2018)*, pages 1–10, New Orleans, USA, 2018.

[30] David Harborth and Sebastian Pape. JonDonym Users' Information Privacy Concerns. In L. Janczewski and M. Kutyłowski, editors, *ICT Systems Security and Privacy Protection - 33rd IFIP TC 11 International Conference, SEC 2018*, pages 170–184, Poznan, Poland, 2018. Springer, Cham.

[31] David Harborth and Sebastian Pape. How Privacy Concerns and Trust and Risk Beliefs Influence Users' Intentions to Use Privacy-Enhancing Technologies - The Case of Tor. In *Hawaii International Conference on System Sciences (HICSS) Proceedings*, Hawaii, US, 2019.

[32] Dominik Herrmann, Jens Lindemann, Ephraim Zimmer, and Hannes Federrath. Anonymity online for everyone: What is missing for zero-effort privacy on the internet? In *International Workshop on Open Problems in Network Security*, pages 82–94. Springer, 2015.

[33] Rob Jansen, Marc Juarez, Rafael Galvez, Tariq Elahi, and Claudia Diaz. Inside Job: Applying Traffic Analysis to Measure Tor from Within. In *Network and Distributed Systems Security (NDSS) Symposium*, pages 1–15, 2018.

[34] Aaron Johnson, Chris Wacek, Rob Jansen, Micah Sherr, and Paul Syverson. Users get routed: Traffic correlation on tor by realistic adversaries. In *Proceedings of the 2013 ACM SIGSAC Conference on Computer & Communications Security*, CCS '13, pages 337–348, New York, NY, USA, 2013. ACM.

[35] JonDos Gmbh. Official Homepage of JonDonym. https://www.anonym-surfen.de, 2018.

[36] Sheharbano Khattak, David Fifield, Sadia Afroz, Mobin Javed, Srikanth Sundaresan, Vern Paxson, Steven J. Murdoch, and Damon McCoy. Do you see what I see?: Differential treatment of anonymous users. In *Network and Distributed System Security Symposium*, 2016.

[37] R. Koch, M. Golling, and G. D. Rodosek. How anonymous is the tor network? a long-term black-box investigation. *Computer*, 49(3):42–49, Mar. 2016.

[38] Ioannis Krontiris, Zinaida Benenson, Anna Girard, Ahmad Sabouri, Kai Rannenberg, and Peter Schoo. Privacy-ABCs as a Case for Studying the Adoption of PETs by Users and Service Providers. In *APF*, pages 104–123, 2015.

[39] Linda Lee, David Fifield, Nathan Malkin, Ganesh Iyer, Serge Egelman, and David Wagner. A Usability Evaluation of Tor Launcher. *Proceedings on Privacy Enhancing Technologies*, (3):90–109, 2017.

[40] Naresh K. Malhotra, Sung S. Kim, and James Agarwal. Internet users' information privacy concerns (IUIPC): The construct, the scale, and a causal model. *Information Systems Research*, 15(4):336–355, 2004.

[41] Naresh K. Malhotra, Sung S. Kim, and Ashutosh Patil. Common Method Variance in IS Research: A Comparison of Alternative Approaches and a Reanalysis of Past Research. *Management Science*, 52(12):1865–1883, 2006.

[42] Akshaya Mani, T. Wilson-Brown, Rob Jansen, Aaron Johnson, and Micah Sherr. Understanding Tor Usage with Privacy-Preserving Measurement. In *2018 Internet Measurement Conference (IMC'18)*, pages 1–13, 2018.

[43] Antonio Montieri, Domenico Ciuonzo, Giuseppe Aceto, and Antonio Pescapé. Anonymity services tor, i2p, jondonym: Classifying in the dark. In *Teletraffic Congress (ITC 29), 2017 29th International*, volume 1, pages 81–89. IEEE, 2017.

[44] Paul A. Pavlou. Consumer Acceptance of Electronic Commerce: Integrating Trust and Risk with the Technology Acceptance Model. *International Journal of Electronic Commerce*, 7(3):101–134, 2003.

[45] paysafecard.com Deutschland. Website paysafecard. https://www.paysafecard.com/de-de/, 2018.

[46] Andreas Pfitzmann and Marit Hansen. A terminology for talking about privacy by data minimization: Anonymity, unlinkability, undetectability, unobservability, pseudonymity, and identity management. 2010.

[47] Philip M Podsakoff, Scott B MacKenzie, J. Y. Lee, and Nathan P Podsakoff. Common method biases in behavioral research: a critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88(5):879–903, 2003.

[48] Christian M. Ringle, S. Wende, and Jan Michael Becker. SmartPLS 3. www.smartpls.com, 2015.

[49] L.D. Rosen, K. Whaling, L.M. Carrier, N.A. Cheever, and J. Rokkum. The Media and Technology Usage and Attitudes Scale: An empirical investigation. *Comput Human Behav.*, 29(6):2501–2511, 2013.

[50] Saad Saleh, Junaid Qadir, and Muhammad U. Ilyas. Shedding light on the dark corners of the internet: A survey of tor research. *Journal of Network and Computer Applications*, 114:1 – 28, 2018.

[51] Carsten Schmitz. LimeSurvey Project Team. http://www.limesurvey.org, 2015.

[52] Blair H. Sheppard, Jon Hartwick, and Paul R. Warshaw. The Theory of Reasoned Action: A Meta-Analysis of Past Research with Recommendations for Modifications and Future Research. *Journal of Consumer Research*, 15(3):325–343, 1988.

[53] Rachee Singh, Rishab Nithyanand, Sadia Afroz, Paul Pearce, Michael Carl Tschantz, Phillipa Gill, and Vern Paxson. Characterizing the nature and dynamics of tor exit blocking. In *26th USENIX Security Symposium (USENIX Security). USENIX Association, Vancouver, BC*, pages 325–341, 2017.

[54] Matthias Söllner, Izak Benbasat, David Gefen, Jan Marco Leimeister, and Paul A. Pavlou. Trust : An MIS Quarterly Research Curation Focus of the Research Curation. *Management Information Systems Quarterly*, (October):1–9, 2016.

[55] Sarah Spiekermann. The Desire for Privacy: Insights into the Views and Nature of the Early Adopters of Privacy Services. *International Journal of Technology and Human Interaction*, 1(1):74–83, 2005.

[56] The Tor Project. https://www.torproject.org, 2018.

[57] Twitter Discussion. Twitter Tweet by Edward Snowden on Tails and Tor. https://twitter.com/Snowden/status/1165297667490103302, 2019.

[58] G.W. van Blarkom, John J. Borking, and J.G.E. Olk. *"PET". Handbook of Privacy and Privacy-Enhancing Technologies.*

2003.

[59] V. Venkatesh and F. D. Davis. A theoretical extension of the technology acceptance model: Four longitudinal Studies. *Management Science*, 46(2):186–205, 2000.

[60] Viswanath Venkatesh, James Thong, and Xin Xu. Consumer Acceptance and User of Information Technology: Extending the Unified Theory of Acceptance and Use of Technology. *MIS Quarterly*, 36(1):157–178, 2012.

All websites have been accessed last on August 25th, 2019.

Distribution Channels of the Tor Online Survey

1. Mailinglists:
   (a) tor-talk[1]
   (b) liberationtech[2]
   (c) IFIP TC 11[3]
   (d) FOSAD[4]
   (e) GI PET[5]
   (f) GI FBSEC[6]
2. Twitter with #tor and #privacy
3. Boards:
   (a) reddit (sub-reddits: r/TOR, r/onions, r/privacy)
   (b) ubuntuusers.de
4. Tor Hidden Service Boards, Sections posted into:
   (a) Darknet Avengers[7], Off Topic
   (b) The Hub[8], Beginners
   (c) Onion Land[9], Off Topic
   (d) 8chan[10], /tech/
   (e) IntelExchange[11], Unverified Users
   (f) Code Green[12], Discussions
   (g) Changolia[13], overchan.random
   (h) Atlayo[14], Posting
5. Personal Announcements at Workshops

1 https://lists.torproject.org/cgi-bin/mailman/listinfo/tor-talk/
2 https://mailman.stanford.edu/mailman/listinfo/liberationtech
3 https://dlist.server.uni-frankfurt.de/mailman/listinfo/ifip-tc11
4 http://www.sti.uniurb.it/events/fosad/
5 http://mail.gi-fb-sicherheit.de/mailman/listinfo/pet
6 http://mail.gi-fb-sicherheit.de/mailman/listinfo/fbsec
7 http://avengersdutyk3xf.onion/
8 http://thehub7xbw4dc5r2.onion
9 http://onionlandbakyt3j.onion
10 http://oxwugzccvk3dk6tj.onion
11 http://rrcc5uuudhh4oz3c.onion
12 http://pyl7a4ccwgpxm6rd.onion
13 http://jewsdid.oniichanylo2tsi4.onion
14 http://atlayofke5rqhsma.onion/