University of Maribor

Faculty of
Criminal Justice and Security

Advances in Cybersecurity 2017

Editors
Igor Bernik
Blaž Markelj
Simon Vrhovec

University of Maribor Press

University of Maribor Press

University of Maribor Press

Advances in Cybersecurity 2017

Editors:
Igor Bernik, Ph.D.
Blaž Markelj, Ph.D.
Simon Vrhovec, Ph.D.

November 2017

Title: Advances in Cybersecurity 2017

Editors: assoc. prof. Igor Bernik, Ph.D. (University of Maribor, Faculty of Criminal Justice and Security),
assist. prof. Blaž Markelj, Ph.D. (University of Maribor, Faculty of Criminal Justice and Security),
assist. prof. Simon Vrhovec, Ph.D. (University of Maribor, Faculty of Criminal Justice and Security).

Reviewers: Igor Bernik, University of Maribor (Slovenia), Luca Caviglione, ISSIA, CNR (Italy), Mehdi Chourib, YAZAKI Europe (United Kingdom), Jan Eloff, University of Pretoria (South Africa), Petra Grd, University of Zagreb (Croatia), Markus Heinrich, Technische Universität Darmstadt (Germany), Stefan Katzenbeisser, Technische Universität Darmstadt (Germany), Jörg Keller, FernUniversitaet in Hagen (Germany), Bradley A. Malin, Vanderbilt University (USA), Blaž Markelj, University of Maribor (Slovenia), Jorge T. Martins, University of Sheffield (United Kingdom), Wojciech Mazurczyk, Warsaw University of Technology (Poland), Fernando Pérez-González, Universidade de Vigo (Spain), Gerardo I. Simari, Universidad Nacional del Sur in Bahía Blanca and CONICET (Argentina), Nicholas Stifter, SBA Research (Austria), Mark Strembeck, WU Wien (Austria), Simon Vrhovec, University of Maribor (Slovenia), Edgar R. Weippl, SBA Research (Austria), Steffen Wendzel, Worms University of Applied Sciences (Germany), Aleš Završnik, Institute of Criminology at the Faculty of Law Ljubljana (Slovenia).

Tehnical editor: Jan Perša (University of Maribor Press).

Cover design: Jan Perša (University of Maribor Press).

All graphic material: Authors of chapters.

Co-published by / Izdajateljica:
University of Maribor, Faculty of Criminal Justice and Security
Kotnikova 8, 1000 Ljubljana, Slovenia
phone +386 1 3000 83 00, fax +386 1 230 26 87
http://www.fvv.um.si, fvv@fvv.uni-mb.si

First published in 2017 by / Založnik:
University of Maribor Press
Slomškov trg 15, 2000 Maribor, Slovenia
phone +386 2 250 42 42, fax +386 2 252 32 45
http://press.um.si, zalozba@um.si

Edition: 1st

Publication type:: e-publication

Available at: http://press.um.si/index.php/ump/catalog/book/296

Published: Maribor, november 2017

CIP - Kataložni zapis o publikaciji
Univerzitetna knjižnica Maribor

004.056(082)

ADVANCED in Cybersecurity 2017 [Elektronski vir] / editors Igor Bernik, Blaž Markelj, Simon Vrhovec. - 1st ed. - El. publikacija. - Maribor : University of Press, 2017

ISBN 978-961-286-114-8
1. Bernik, Igor, 1985-
COBISS.SI-ID 93431041

ISBN: 978-961-286-114-8 (PDF)

DOI: https://doi.org/10.18690/978-961-286-114-8

Price: Free copy

For publisher: full prof. Igor Tičar, Ph.D., Rector (University of Maribor)

DOI https://doi.org/10.18690/978-961-286-114-8
ISBN 978-961-286-114-8
© 2017 University of Maribor Press
Available at: http://press.um.si

University of Maribor Press

Advances in Cybersecurity 2017

IGOR BERNIK, BLAŽ MARKELJ & SIMON VRHOVEC

Abstract Understanding the cyberspace and awareness of its effects impacts the lives of all individuals. Thus, the knowledge of cybersecurity in both organizations and private operations is essential. Research on various aspects of cybersecurity is crucial for achieving adequate levels of cybersecurity. The content of this scientific monography provides answers to various topical questions from the organizational, individual, sociological, technical and legal aspects of security in the cyberspace. The papers in the monography combine the findings of researchers from different subareas of cybersecurity, show the effects of adequate levels of cybersecurity on the operations of organizations and individuals, and present the latest methods to defend against threats in the cyberspace from technical, organizational and security aspects.

Keywords: • Cybersecurity • cyber resilience • mobile security • digital privacy • IoT security •

CORRESPONDENCE ADDRESS: Igor Bernik, Ph.D., Associate Professor, University of Maribor, Faculty of Criminal Justice and Security, Kotnikova 8, 1000 Ljubljana, Slovenia, e-mail: igor.bernik@fvv.uni-mb.si. Blaž Markelj, Ph.D., Assistant Professor, University of Maribor, Faculty of Criminal Justice and Security, Kotnikova 8, 1000 Ljubljana, Slovenia, e-mail: blaz.markelj@fvv.uni-mb.si. Simon Vrhovec, Ph.D., Assistant Professor, University of Maribor, Faculty of Criminal Justice and Security, Kotnikova 8, 1000 Ljubljana, Slovenia, e-mail: simon.vrhovec@fvv.uni-mb.si.

DOI https://doi.org/10.18690/978-961-286-114-8 ISBN 978-961-286-114-8

Available at: http://press.um.si.

Table of Contents

Introduction

IGOR BERNIK, BLAŽ MARKELJ & SIMON VRHOVEC

In the light of the recent mega breaches, information security has proved to be an indispensable building block of corporate security and an essential element for providing overall economic prosperity. Although there is a broad range of guidelines on how to manage information risks adequately, they still seem to fail to adequately address all the weak points and information security aspects in practice. This scientific monography aims to advance the knowledge in various cybersecurity aspects and their inter-connections from real-life applications.

The first chapter addresses the feasibility issue of best practices in information security in order to showcase them as ideal-type situations. The grounds for this concern were established through the testing of the information security performance model (ISP 10×10M). The model was developed in cooperation with security experts and represents a performance evaluation framework designed primarily for small and medium-sized businesses (SMBs).

For ensuring information security of an organization, it is important that all its members accessing the cyberspace recognize and deal with cyber threats individually. Even though appropriate technical cybersecurity measures are commonly employed in organizations, its members are still directly exposed to cyber threats and are one of the key weak points of ensuring information security of the whole organization. Omission of adequate cybersecurity measures by individuals can be at least embarrassing for the organization and can even critically affect its business in some cases, e.g., through loss of trust. In the second chapter a new model based on the insights from extant research in various fields by combining protection motivation theory, technology threat avoidance theory, and psychological reactance theory into a unified model is proposed and tested.

Concept drift analysis propose the pre-processing step dedicated for anomaly detection systems to counter cyberattacks. Such approach could be a step towards lifelong learning intelligent cybersecurity system. Such system could use the previously learnt knowledge to properly detect cyberattacks in the ever-changing networked environments without the necessity to re-learn from scratch. The third chapter explores how could such approach improve the detection rate e.g. of the obfuscated SQL injection attacks and decrease the false positive rates by properly adjusting to the concept drift in the data.

The fourth chapter explores the cybersecurity aspect of nuclear facilities which are a part of a national critical infrastructure that is completely dependent on information technologies. Traditionally, these systems were completely isolated from external networks, but recent transition to digital technology has changed the nature of these systems thus enabling extensive interconnections. Therefore, a great amount of Industrial

Control Systems (ICS) are connected to the cyber space over corporate networks. Wireless technologies and remote administration are also widely used. The ICS are no longer air-gapped as they used to be and consequently much more vulnerable to external threats.

The fifth chapter explores the use of mobile devices in hospitals as one of the key components of the healthcare critical infrastructure. Health care professionals are increasingly using mobile devices in their everyday work to improve patient care. Hospitals may however fail to adequately address the use of mobile devices and adapt their information security policies in time. Health care professionals may use both their personal and work mobile devices for their everyday work. Sometimes they do it without adhering to an adequate hospital information security policy. Adhering to information security policy is positively correlated with perceived data breach consequences for both the patients and the hospital.

The sixth chapter reports on a study among two generations, i.e., students and employees, in order to generate a comprehensive overview of their use of mobile devices and observing the inter-generational differences. The results of the study also served as a basis for determining whether users of mobile devices are at risk of becoming victims of information security threats. Results demonstrate that the awareness of threats arising from the use of mobile devices is higher among the employees. The lack of knowledge and rules concerning the use of mobile devices represents the main problem as the victims of mobile devices information security threats have various possibilities at their disposal to protect themselves.

Malware has become more harmful than in the past as the number of intelligent systems and Internet-connected devices increased dramatically. Mobile device has become the predominant means of accessing personalized computing services such as email, banking, etc. However, this rapid deployment and extensive availability of mobile apps has made them attractive targets for various malware. Therefore, one of the most important issues in cybersecurity has become the detection of previously unknown malware in the shortest time possible in order to stop it from becoming common hazards and to harm users. Antimalware systems detect existing malware successfully but they are far from achieving the same detection performance for unknown malware (zero-day malware). For this purpose, machine learning methods were applied to detect and classify malware in the seventh chapter. But methods are commonly based on information gathered via dynamic analysis to achieve better accuracy with machine learning based techniques. A novel model based on deep learning for the prediction of mobile malware without requiring execution in an isolated environment is proposed.

Although conceptually not new, ransomware recently regained attraction in the cybersecurity community: notorious attacks in fact have caused serious damage, proving their disruptive effect. This is likely just the beginning of a new era. According to a recent intelligence report by Cybersecurity Ventures, the total cost due to ransomware attacks is predicted to exceed $ billion. How can this disruptive threat can be contained? In the eight chapter, the future of ransomware is discussed. Current anti-ransomware solutions are effective only against existing threats, and the worst is yet to come. Cyber criminals will design and deploy more sophisticated strategies, overcoming current defences and, as it commonly happens in security, defenders and attackers will embrace a competition that will never end. In this arm race, anticipating how current ransomware will evolve may help at least being prepared for some future damage. Discussing how current ransomware could become even more disruptive and elusive is crucial to conceive more solid defence and systems that can mitigate zero-day ransomware, yielding higher security levels for information systems, including critical infrastructures such as intelligent transportation networks and health institutions.

Video surveillance can be defined as using video cameras to observe an area. These systems can have many benefits, however there are also many risks. violation of privacy is most often mentioned as a problem. One of the most vulnerable groups whose privacy needs to be protected are children. The ninth chapter analyses the importance of children privacy protection in video surveillance and proposes a model for children privacy protection based on age estimation. The proposed model uses automatic age estimation in order to distinguish between children and adults. Age can be estimated in different ways, the model proposed in this chapter classifies people by using their face anthropometry.

The trend of digitalization fostered by Internet of Things technologies is characterized by a huge potential for innovations on one side and security and privacy threats on the other side. The emerging IoT cloud services other great opportunities, since the complexity of providing IoT services is significantly reduced. On the other hand, the level of security and privacy is not transparent to the users of the service in addition to the provided documentation. In the tenth chapter, the possibilities of testing security and privacy requirements of IoT cloud services from a user perspective based on a generalized architecture of IoT cloud services are investigated. The proposed test framework is used to evaluate the IoT cloud services of the providers Amazon, Microsoft, Google and ThingSpeak.

Technology revolution seems to be unstoppable and soon we will live in a world in which all devices will be connected to the Internet. The Internet of Things (IoT) is boosting due to the reduction of size and price of the sensors. At home, we can find dozens of gadgets from smart cars that can park themselves to fridges which know what kind of products are being stored inside and can determine whenever a food item needs to be replenished. However, the problem is that many of these devices are being made the same way as 20 years ago without focusing on the security issues. The eleventh chapter reviews the vulnerabilities and security problems that the different devices may have and a model to estimate the risk score of them is proposed. The following variables were used to calculate the score: authentication type, default credentials, vulnerabilities, hacking news and number of exposed devices to the Internet (more than 20,000 discovered). Then these scores were combined to find out how secure (or insecure) your home is, based on their individual values. The model is used by a web application that lets the users select the IoT devices they have at home and returns the security risk of suffering an attack.

In the twelfth chapter, Cycle Structure and Reachability Analysis for Cipher Spritz with small N where Spritz has been proposed as a replacement for RC4. For embedded applications that use it as a stream cipher or pseudo-random number generator with smaller parameter N than the standard N = 256, the choice of N should be as small as possible for performance, but large enough to provide sufficient security. Hence, we investigate which fraction of the state space is reachable with key setup and subsequent output, and which cycle length can be expected for small N. Next to some elementary theoretical investigations, we experimentally do an exhaustive search on the state space for N = 4,6,8.

The future is here, and it requires that the participation of different actors, professionals, and especially all interest groups be already included in the plans for the development of cybersecurity strategies today.

University of Maribor Press

Information Security Management Practices: Expectations and Reality

KAJA PRISLAN, BRANKO LOBNIKAR & IGOR BERNIK

Abstract Information security has proved to be an indispensable building block of corporate security. Although there is a broad range of guidelines on how to manage information risks adequately, they fail to manifest their value in practice. This paper aims to address the feasibility issue of best practices in information security, in order to showcase them as ideal-type situations. The grounds for this concern were established through the testing of the information security performance model (ISP 10×10M). The model was developed in cooperation with security experts and represents a performance evaluation framework designed primarily for small and medium-sized businesses (SMBs). When the model was implemented in 20 Slovene SMBs, results showed that their strategies differed from professional recommendations. Factors that ranked high on the experts' list were mainly considered as the lowest priority by practitioners. We argue that the actual state-of-play deviates from professional recommendations. The lack of knowledge and skilled security staff were most likely the main sources of wrongly set priorities.

Keywords: • Information security • management • best practices • performance • state-of-play •

CORRESPONDENCE ADDRESS: Kaja Prislan, Ph.D., Assistant Professor, University of Maribor, Faculty of Criminal Justice and Security, Kotnikova 8, 1000 Ljubljana, Slovenia, e-mail: kaja.prislan@fvv.uni-mb.si. Branko Lobnikar, Ph.D., Associate Professor, University of Maribor, Faculty of Criminal Justice and Security, Kotnikova 8, 1000 Ljubljana, Slovenia, e-mail: branko.lobnikar@fvv.uni-mb.si. Igor Bernik, Ph.D., Associate Professor, University of Maribor, Faculty of Criminal Justice and Security, Kotnikova 8, 1000 Ljubljana, Slovenia, e-mail: igor.bernik@fvv.uni-mb.si.

DOI https://doi.org/10.18690/978-961-286-114-8.1 ISBN 978-961-286-114-8

Available at: http://press.um.si.

1 Introduction

In the past decade, we have witnessed a fast-paced development of cybercrime, which has in fact become one of the top security threats for organisations. As a result, information security gained recognition for having a high business value and a critical impact on overall economy. However, in the era of information power, highly motivated attackers and zero-day vulnerabilities, information assurance and cyber resilience are not easy to achieve.

The effectiveness of information security depends on the application of different measures, which must be implemented in adequate proportion and balance [1], [2]. Experts working in this field agree that successful information risk management exceeds situational (technical) measures and that information security is in fact a managerial issue. Moreover, managers are the main agents and leading factors of information security performance [ISP] [3], [4]. Accordingly, questions, such as how to approach security decision-making, organise leadership and integrate security into everyday business, are being actively investigated [3], [5]. An overview of current organisational practices shows that the practitioners' awareness of the value of information security has improved in recent years, however, overall competences have been worse than anticipated [6], [7]. Numerous reports [8]–[11] highlight that organisations are not effective in managing information security threats and that only few comply with established guidelines. Considering that most information incidents could be avoided by introducing basic measures and that detection capabilities are underdeveloped, the current state-of-play suggests that there is a lack of coordination between entities responsible for security. Security practices are not able to keep up with the developing knowledge and recommendations produced by security professionals and researchers.

One might argue that organisations simply do not possess resources for a more proactive development of information security and the implementation of professional recommendations. For many years now, financial aspects (insufficient funding or budget justification) have in fact been emphasised as a major obstacle to ISP [12], [13], which is a plausible argument explaining the underdeveloped practice to some degree, but performance cannot depend solely on budgets. In this case, security would be a luxury available only to the few, whereas in realty it is a basic need that simply must be fulfilled. Furthermore, many guidelines in the field of information security management [ISM] that provide recommendations for higher cost-effectiveness, easier planning and decision-making are available.

That said, there is an obvious gap between theory and practice [14], and there are two possibilities why the available knowledge, best practices and know-how are not manifested in reality. Either managers are not qualified or security recommendations and professional expectations are not reasonable. The main objective of this paper is to provide some answers to these dilemmas and raise questions regarding this gap.

2 Information security: state-of-play and state-of –knowledge

For several years now, organisational (information) security capabilities have been reported as inefficient and underdeveloped. Experts are persistent in their warnings stating that the overall ISP is slowly decreasing [15], while recognising the fact that constant changes in the threat environment are difficult to follow. As a result, many businesses, corporations and important public services experience major disruptions, data loss or high-risk exposure on a daily basis. However, reports show that the overall materialised threats fall into basic and well-known patterns [16]–[18]. Despite the available knowledge in the (information security) risk management domain, we remain practically unable to control basic threats, let alone more organised and sophisticated ones.

The contemporary issues of information security are portrayed in different state-of-play reports. For example, the likelihood of an information incident on a yearly basis is extremely high for businesses (it ranges between 70-90%) [19]–[21]. Furthermore, the majority of victimised companies struggle to rebound after information breaches and cyber attacks. The overall detection and recovery take days or even months, while many also experience a drop in their share value after a public disclosure of information incidents [22], [23].

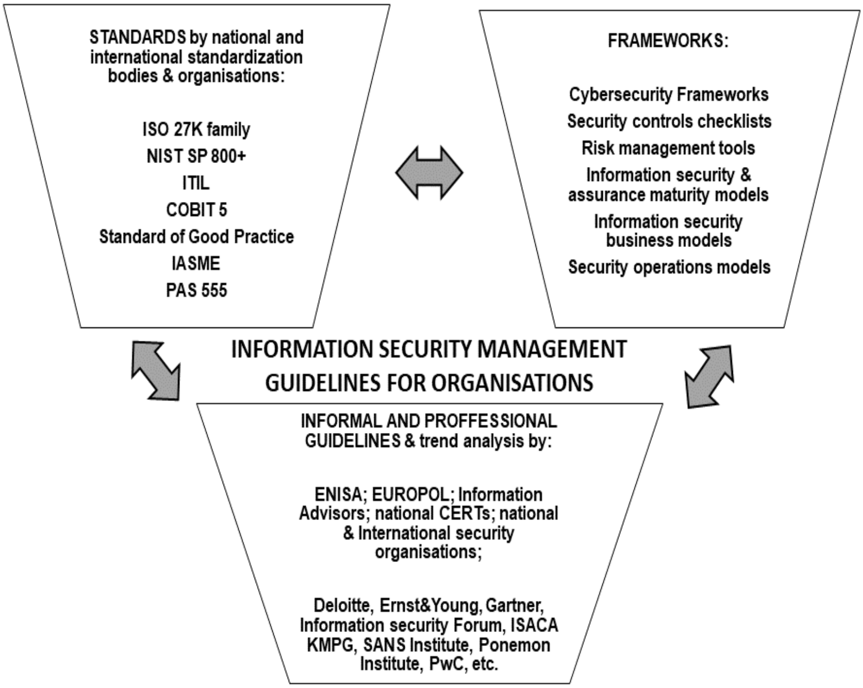

STANDARDS by national and international standardization bodies & organisations:
ISO 27K family
NIST SP 800+
ITIL
COBIT 5
Standard of Good Practice
IASME
PAS 555
FRAMEWORKS:
Cybersecurity Frameworks
Security controls checklists
Risk management tools
Information security & assurance maturity models
Information security business models
Security operations models
INFORMATION SECURITY MANAGEMENT GUIDELINES FOR ORGANISATIONS
INFORMAL AND PROFFESSIONAL GUIDELINES & trend analysis by:
ENISA; EUROPOL; Information Advisors; national CERTs; national & International security organisations;
Deloitte, Ernst&Young, Gartner, Information security Forum, ISACA KMPG, SANS Institute, Ponemon Institute, PwC, etc.

Figure 1: Types of ISM guidelines

The latest self-reports from Chief Information Security Officers (CIOs) demonstrate that practitioners are not confident about their security situation and believe that the current security controls are failing to protect their business [24], [25]. Similarly, a report about the maturity of information security shows that the majority of companies do not develop or operate with basic information security controls and processes (e.g. threat assessment, vulnerability assessment, monitoring tools and processes) [26]. It is also important to note that smaller companies are the most vulnerable part of the business ecosystem [27], [28]. They are in fact just as much of a target than larger businesses, but they often believe that they are not vulnerable to cybercrime [29], [30]. In reality, attackers prefer to focus their efforts on businesses with less security resources, which is why SMBs are becoming prime targets of cyber criminals. Not only do SMBs have large amounts of valuable data and low security levels (which makes them vulnerable to basic automated attacks), but they also represent an entry point for larger companies [18], [31].

Such a situation shows that the much needed transition of the information security function to the strategic level is stalled [32], [33]. To help practitioners gain a sense of direction in the field of ISM, numerous types of guidelines (standards, models and frameworks) from the research and professional opus are available. Some references for further investigations in all three domains are presented in Fig. 1.

The ISO ISMS 27k family of standards (including ISO 22301 – business continuity) is the most recognised and widely accepted set of standards, where the ISO/IEC 27001 (and its accompanying 27002 standard – code of practice) represents the leading specification for governing information security in organisations, followed by the ISO/IEC 27032 (cybersecurity), ISO/IEC 27035 (incident management) and ISO/IEC 27005 (risk management). Apart from the ISO ISMS, other guidelines and publications, such as NIST (Cybersecurity Framework and Special publications 800+), ITIL (Axelos), COBIT 5 (ISACA) and the Standard of Good Practice (ISF), are also highly recognised. These standards have a broad application, but they can generate high financial costs, especially when considering their certification. In addition, informal guidelines and trend reports are regularly provided by different national, international, non-profit and analytic organisations specialised in the security domain.

Assessing security needs and compliance of security measures is one of the most important and basic steps, especially in terms of governance [34]. Without such an assessment, the determination of controls adapted to organisational needs and strategies proves extremely difficult. For this purpose, many security frameworks are also available for organisations to evaluate their security needs and state-of-play /performance.

Considering all available publications, the lack of knowledge should not be an issue for practitioners. Several studies confirm that compliance with standards may have a positive impact on ISP and organisations' resilience [35], [36]. On the other hand, companies support compliance in practice only to some extent, whereas regulatory norms are their main objective. According to studies, organisations consistently follow regulations, while only a small share is actively engaged in standardisation and accreditation [37]. The reason for such a distribution lies in the fact that most managers consider regulatory compliance as the most effective mean to justify security funding [38]. This means that compliance is seen as a collateral to conformity with law and obligatory directives. However, if there is no sign of reduction in the number or the severity of reported breaches, the effectiveness of ongoing information security practices should be called into question [39].

It is possible to argue that smaller or understaffed companies find it difficult to tackle the abundance of guidelines, standards and requirements [40], [41]. A recent global CEO survey [42] found that 80% of top management see over-regulation as one of the main threats to their organisation's growth objectives. Ernst&Young [10] reports that 23% of organisations believe that their effectiveness is lower due to the fragmentation of regulations. Policy and standards implementation are among the least developed processes and only 10% of respondents believe that their organisation has a mature position in this area. One of the major concerns, possibly connected to the latter issue, is a severe shortage of cybersecurity talent on a global scale, since many organisations struggle with the lack of security skills [25], [43]. It would also be reasonable to assume that the low application of standards in practice relates to their generic nature. As many experts already pointed out, the problem is that guidelines are too common and untailored to the management needs and expectations [44]–[46]. This is why we can reasonably assume that the disparity of publications, the neglect of industry specific demands and the lack of interest to unify some of the most common operations, are correlated to the above described problems stemming from defective information security practices.

Hence, the information security guidance for industry requires certain changes. Currently, the situation is unmanageable for most organisations. The number of standards is inconclusive, while most of them contain an excessive number of recommendations that are difficult to realise; at the same time, the process of standardization is overly complex and time consuming [47]. Moreover, these standards usually presume certain ideological conditions and neglect actual circumstances, organisational needs, restrictions and practical constraints [46], [48]. Standards should not hinder or confuse practitioners, but reflect best practices and present recommendations that are achievable, effective and helpful.

At this point, it should be emphasised that our intention is not to oppose existing research endeavours and studies regarding the link between standards' implementation and organisational performance. We believe that a systematic and well-organised standardisation, which follows a carefully selected publication, can indeed improve the overall information security position. There are three issues we wish to analyse in this paper: 1) to define the level of agreement regarding security priorities among experts and practitioners; 2) to identify the level of organisational compliance with security standards; and 3) to analyse information security strategies. Furthermore, we aim to define common views about ISP requirements with the purpose of facilitating convergence between different security entities. Our goal is to provide feed-back for professionals and promote the drafting and modelling of guidelines that are tailored to actual organisational needs.

3 Information security performance - research

In order to investigate the current industry information security situation, we conducted a research study among information security experts and practitioners in Slovenia regarding their opinions and practices. The presented research study was conducted as part of a larger project (i.e. Measuring information security in organisations) that consisted of four consecutive phases: 1) firstly, an overview of established information security standards was conducted and 2) secondly, a model (i.e. ISP 10×10M) for SMBs was developed in cooperation with experts. The model was then 3) applied to a small sample of SMBs, whereas 4) a comparison between professionals' and practitioners' reports was finally conducted. The core idea of this project was that smaller companies need more practical and problem-oriented information security solutions for management.

Although the research project had multiple objectives, this paper only presents partial findings related to the subject matter regarding the practicality and relevance of existing information security recommendations. The focus is therefore on the fourth phase of the project; however, we will summarise the background and related segments of other phases with a view of providing a better understanding of the project methodology. More details about the process of the ISP 10×10M development and its structure are provided in an earlier publication by Bernik & Prislan [49].

The model itself is composed of measures that are commonly recommended and included in established ISM publications. The analysis included many different guidelines, however, most of the reviewed findings and measures derived from international standards,[1] and trend analysis reports.

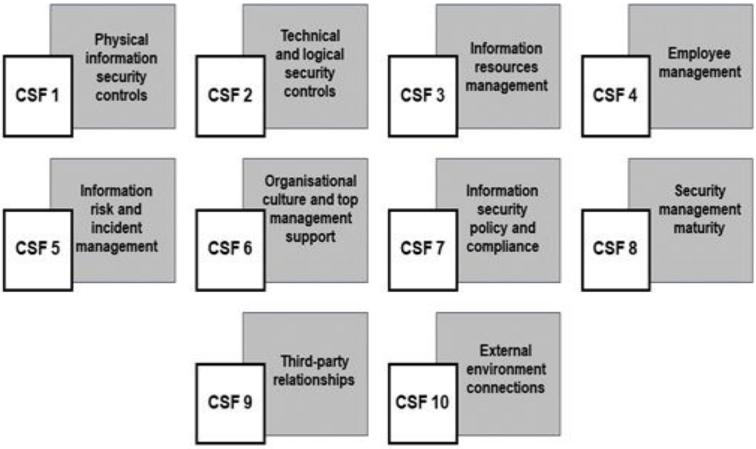

CSF 1
Physical information security controls
CSF 2
Technical and logical security controls
CSF 3
Information resources management
CSF 4
Employee management
CSF 5
Information risk and incident management
CSF 6
Organisational culture and top management support
CSF 7
Information security policy and compliance
CSF 8
Security management maturity
CSF 9
Third-party relationships
CSF 10
External environment connections

Figure 2: Critical success factors of the ISP 10×10M

The ISP 10×10M is constructed of 10 factors (i.e. CSF, critical success factors), whereas each one of them further consist of 10 specific CSF related measures (i.e. KPI, key performance indicators). The model consists of 100 information security controls in total. The CSFs' of the model are presented in Fig 2. The model is structured in a way that provides SMBs with a practical solution for assessing their information security performance. When the model was implemented to the target group of Slovene organisations, the results showed that their information security is generally defined and managed, however, the security areas and measures are prioritised differently to what experts would expect. The discrepancies between groups and possible arguments for practical deficiencies are presented in the next section. In order to provide a coherent comparison between studies, we first introduce the methodology applied for analysing the perceptions of experts and practitioners.

3.1 Sample and methodology

Since different factors and indicators have a different impact on the final information security situation, we had to determine the extent to which the measures included in the model were generally relevant for good practice. In order to provide objective weights, the indicators were evaluated in cooperation with professionals dealing with information security on a daily basis. With the collected data, we were able to weight and categorise individual indicators, establish a structure of the model, identify categories of correlated controls, develop a method for calculating the ISP score and define critical performance levels. The model was then practically tested on and applied to a sample of medium-sized private sector companies. Prior consent of the management was obtained from all participating entities for conducting the research.

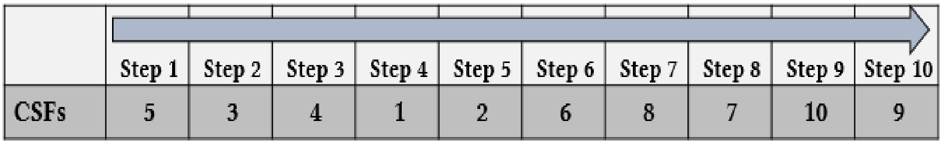
In the first research study, we had a medium-sized sample of 43 IT security experts, mainly from computer, telecommunications and financial organisations, with higher education qualifications and private sector experience. Experts provided their opinion (grades) about the importance of CSFs and KPIs[2] [49]. In the second research study, we had a small sample of 20 Slovene businesses, categorised into four business sectors: production & industry (40%), retail (20%), energy (25%) and services (15%). Most participating organisations are present on global markets (75%) and employ from 50 to 149 (60%) or up to a maximum of 250 employees. The majority of respondents reported high levels of informatisation and information risks.

Participants from the first research study were asked to evaluate the importance of individual criteria for an efficient implementation of information security. When assigning the level of importance, value 1 meant that an indicator was less important, while value 5 meant that an indicator was of critical importance for SMBs. The same method was used in the second research study, whereby organisations had to evaluate and grade the same criteria based on their security priorities and state-of-play, where 1 meant that the organisation does not comply with the criteria and 5 represented a state of full compliance. The main difference here was that experts gave an opinion about the importance of criteria, while organisations evaluated the criteria based on their practical situation.

3.2 Results

Before we address the issue regarding organisational deviations, we ought to present the results of the first research study. Experts' point-of-view must be understood in order to discuss the results of the second research study and conduct their comparison.

The most favourable business (information security) plan for SMB's according to experts' recommendations is presented below. The factors are organised by priority according to the evaluations provided in the model assessment. Similarly, the KPIs within individual factors were ranked from the most to the least important, which was significant for further model development.



	Step 1	Step 2	Step 3	Step 4	Step 5	Step 6	Step 7	Step 8	Step 9	Step 10
CSFs	5	3	4	1	2	6	8	7	10	9

Figure 3: ISP recommended factor prioritisation

Fig. 3 demonstrates that information risk management (CSF 5) ranks first and is therefore the most important and influential CSF of the model, where measures (KPIs), such as alternative locations and hot sites; business continuity plans and early warning systems (such as IDS, IPS, SIEM), were identified as the most critical. These are followed by other significant criteria that fall under the scope of CFS 3 and 4 (information sources and user management), for example archives and back-ups, personal data protection, user rights control, training of security staff, employees' awareness and remote access security protocols. Technical, logical and physical controls (CSF 1 and 2) are also rated highly, as was to be expected, since they generally represent first line of defense and are therefore one of the first steps organisations (should) take when planning and establishing their information security. Experts participating in the research evaluated the compliance of information security (CSF 7) as a less important factor. They believe that the fulfilment of formal conditions does not have as strong an impact as, for instance, technical and operational controls. Factors related to the control of external environment and relations with third parties (CSF 10 and 9) were also graded with lower values, which is reasonable, since these external elements are the areas which are more difficult to manage. Furthermore, organisational culture and security management maturity (CSF 6 and 8) are not considered as top priorities in our model. Even though the management of social and psychological aspects are in fact important, the high level of ISP is predominated by high-quality technical controls. The aforementioned management processes are of more strategic nature and require an operational baseline.

Provided with such a layout, organisations ought to focus on technical and situational measures to cope with critical risks and vulnerabilities in the initial phases, while measures in the subsequent phases should gradually evolve from the operational to the strategic level. Security mechanisms should be more targeted, specific and complex with every subsequent level.

The results of the first research study were then coupled with the findings of the second study, where we focused on the identification of priorities and practices in organisations in order to determine whether they followed the experts' recommendations. We also aimed to evaluate their overall performance according to the model predispositions. The data collected were first analysed with the basic descriptive statistics and then multiplied with the designated variable weights. The result of information security performance of an organisation is obtained by calculating the sum ($W_{1\text{-}100} \times G_{1\text{-}100}$), where G represents a grade of KPI development (on a scale from 1 to 5) provided by the organisation. The weights of KPIs were determined in the first research study on the basis of experts' evaluation. Each indicator (1,12-1,024) and factor (0,6-1,2) was given a unique weight and had a different impact on the outcome – the final KPI weight (W) was then calculated by multiplying the KPI's initial weight (IW) and the CSF weight (FW). The normalisation of results was avoided by defining such a scale of weights that provide a score between 20 (min.) and 100 (max.) when multiplied with individual grades.

Table 1: Comparison of results between studies

CSF	Group (E/O)	Mean	Score (pt)	Gap (E/O ratio)	t-test (2-independent samples)			
					t	Df	Sig.	MD
1	Experts	4.05	11	0.775	0.82	57	0.41	0.21
	Org.	3.85	8.53					
2.	Experts	4	10.5	0.766	0.97	60	0.33	0.2
	Org.	3.8	8.04					
3.	Experts	4.23	13	0.747	2.53	58	0.01***	0.55
	Org.	3.68	9.71					
4.	Experts	4.1	12	0.742	2.27	53.7	0.03**	0.44
	Org.	3.66	8.9					
5.	Experts	4.24	14	0.565	6.51	60	0.00***	1.5
	Org.	2.74	7.91					
6.	Experts	3.71	9.5	0.771	-0.92	56	0.36	-0.2
	Org.	3.91	7.32					
7.	Experts	3,56	8	0.754	-0.81	57.46	0.42	-0.16
	Org.	3.72	6.03					
8.	Experts	3.57	9	0.63	1.93	60	0.06*	0.45
	Org.	3.13	5.67					
9.	Experts	3.38	6	0.778	-2.11	59.97	0.04**	-0.43
	Org.	3.81	4.67					
10.	Experts	3.43	7	0.684	0.1	57.23	0.92	0.02
	Org.	3.41	4.79					
	Average score (O): points			71.65				

***p-value <0.01; **p-value <0.05; *p-value <0.1

Table 1 summarises the mean values of ten factors from both research studies,[3] as well as weighted results (i.e. scores). This is followed by a calculation of the gap — the ratio between the maximum possible result of the factor with respect to the average points collected by organisations.[4] By comparing these results, we were able to draw parallels between findings and determine whether organisations agree with experts regarding the prioritisation of security areas and therefore design their security plans accordingly to meet the recommendations.

T-tests, which are used to compare means between groups and test their statistically significant differences, were conducted on non-weighted average values. The report shows differences between experts' opinion and practitioners' experience. Statistically significant differences are evident in CSFs 3, 4, 5 and 9, while deviations are also notable in CSF 8.[5] When analysing the means of CFSs, the entities participating in the second research study did not meet the recommendations of experts in all cases, except in the ninth factor, which exceeded expectations. The gap ratio points to similar conclusions, however, it provides a more detailed insight into the accuracy of organisations' priorities and plans. For example, t-tests only show the difference between the perceived relevance of a given area between groups, while the gap addresses weighted results and reports if the security measures of a given CSF are addressed in a proper order. If we focus on CSF 10, we can observe that the mean values between groups are not particularly different, however, the gap ratio is low. This indicates that the efforts and investments dedicated to the related measures are not accurate according to experts, even though organisations share the same opinion about the (low) factor relevance. A low weighted result (i.e. large gap) reports that less important KPIs are highly valued in practice, while most important KPIs are lower on the priority list.

According to organisations' assessment of all ten factors, CSF 6 (Organisational culture and management support) ranked first and therefore represents the most developed information security area in practice. However, a more detailed analysis of the factor shows that leadership participation and support is almost non-existent. In terms of the overall results, CSF 1 (Physical information security controls) and CSF 9 (Third-party relationships) were also amongst top priorities on practitioners' list. On the other hand, CSF 5 (Information risk and incident management) is the worst-rated factor, which is contradictory, since experts consider it to have the highest impact on information security efficiency. The category of least developed factors includes CSF 8 and 10 as well, which are related to security management competences and control over the external environment. The underdeveloped security management function is actually in line with a low level of leadership support, which can in fact represent a major barrier for the development of an organisation. Given the role of the management, its absence could be a possible cause of the poor situation regarding risk management related activities. Low results of the tenth factor have indeed been anticipated, since it consists of measures that organisations naturally develop at later stages.

If we now turn to the overall results, the analysis shows that organisational practices could be evaluated as intermediate or mid-stage. The average score achieved by organisations represents 71.65 points. When an organisation evaluates its capabilities in 100 areas, the model provides an answer about its information security performance with a score, which is followed by a categorisation into one of the six information security levels. According to the results obtained in the second research study, the majority of organisations ranked between the third and fourth levels of the model. However, the deviations between participating entities were relatively high in both positive and negative directions. By analysing the collected data, we were additionally able to categorise information security priorities according to individual factors of both groups. A comparison of results obtained in both studies with regard to the recommended and the actual order of CSFs are presented below.

Fig. 4, which is based on non-weighted results, shows: 1) the sequence of areas and measures that organisations should develop according to the opinion of experts; 2) the actual priorities of organisations; and 3) the extent to which these priorities are in line with recommendations. Factors in the first column reflect experts' recommendations and are ranked from the most important to the least important; in the second column, they are categorised from the most to the least developed in practice, while the attached icons contain suggestions for future development of these areas.

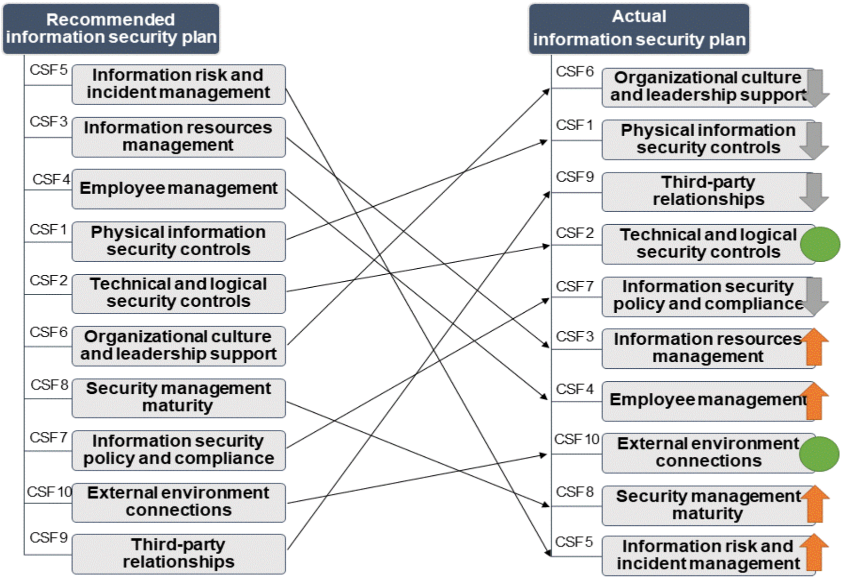

Recommended information security plan
CSF 5 Information risk and incident management
CSF 3 Information resources management
CSF 4 Employee management
CSF 1 Physical information security controls
CSF 2 Technical and logical security controls
CSF 6 Organizational culture and leadership support
CSF 8 Security management maturity
CSF 7 Information security policy and compliance
CSF 10 External environment connections
CSF 9 Third-party relationships
Actual information security plan
CSF 6 Organizational culture and leadership support
CSF 1 Physical information security controls
CSF 9 Third-party relationships
CSF 2 Technical and logical security controls
CSF 7 Information security policy and compliance
CSF 3 Information resources management
CSF 4 Employee management
CSF 10 External environment connections
CSF 8 Security management maturity
CSF 5 Information risk and incident management

Figure 4: ISP in theory and practice

As we can see from the second column, participating organisations somehow comply with the professional recommendations in the areas of CSF 2 and 10, which are related to technical controls and positioning in the external environment. On the other hand, the most substantial negative deviations are again observed between CSF 5, 8 and 3, while positive differentiations are evident in CSF 1, 6, 9 and 7. These findings point to a problematic situation, since negative deviations occur precisely in those areas that experts consider as the most significant, while those areas (for example CSF 9 and 6), which are less influential, are practically most regulated. This also means that entities with low results can improve their ISP relatively quickly, since the improvement of unregulated measures would greatly contribute to the final result. Given the observed deviations, we can deduce that organisations only act efficiently with respect to two factors, while they should reorganise their priorities to include other factors. In areas that appear most developed, it would be advisable to reduce investments slightly and reallocate them to lower ranked factors, as they have a greater impact on information security performance. In general, the results strongly confirm our original arguments and assumptions regarding underdeveloped and somehow misaligned organisational practices.

4 Discussion

This paper presents a comparative study, the main objective of which was to analyse the disparities between theory and practice in the area of ISM. Literature review shows that many existing professional recommendations are not tailored to those organisations that represent the most vulnerable part of the cyber-business ecosystem. The situational overview also implies that practical activities do not follow professional guidelines, thus creating a gap between theory and practice. However, the problem was not yet closely studied.

For this purpose, we initiated a research project aimed at building an information security performance model for SMBs, which is based on a systematic review of standards and trend reports. Our goal was successfully achieved and presented by Bernik & Prislan [49], whereas parts of continuing work in the project are presented in this paper. Here, the model represents the framework for our research studies in which we investigated the importance of 100 security controls for an effective information security. The comparison was conducted between experts' opinions and reports of participating organisations regarding compliance with these measures. On this basis, we were able to identify (information security) priorities as defined by professionals and, conversely, as performed by organisations. The purpose of this exercise was to determine whether there were any differences between the positions of the two groups.

The first research study, which was based on ISM standards review and supported by experts, showed that measures related to risk and incident management, employee and user control, information analysis, technical and physical protection have the highest impact on ISP. On the contrary, the second research study conducted among business entities (i.e. practitioners) gave somewhat opposite findings regarding their practical activities. However, following the implementation of the model, the final performance results were neither extremely low nor excellent. The practical situation shows that the state-of-play in the selected Slovene companies could be evaluated as intermediate on average; a further categorisation of individual entities and their compliance with controls indicates that the majority maintains a reactive position. Approximately 50% of all measures presented in the model are implemented appropriately and sufficiently. The most important observation to emerge from the data comparison is that the situation in practice differs from the recommended strategy. The results showed that the information security priorities of most participating organisations do not follow experts' guidelines regarding efficiency and high performance.

The main conclusion is that activities related to management, information risk, information sources and user control are less developed than is recommended and should be improved in the future. Amongst all findings, the realisation that the worst performance is related to the most important factors appears to be the most outstanding one. Although the situation in organisations cannot be considered alarming following the application of the model, we find that there is a significant gap that indicates a deficient and inadequate state-of-play. The deviations are the highest in peripheral factors, i.e. those that are supposed to be the most and the least important. Too much attention is dedicated to less relevant areas, while important ones are mostly ignored.

There are many possible explanations why such deviances occur in practice. Most probably, practical inconsistencies result from a combination of different causes. Hence, we propose to explore some of the most plausible arguments. One of the possible reasons for misaligned management practices could be related to the unsystematical employment of information technology without considering the encompassing vulnerabilities and security impacts. It seems that companies would rather take the instrumental approach to learning, whereby their actions are related to the experience or factual triggers (i.e. threats). We believe that this could be related to the overconfidence effect and the traditional mindset governing the security function, which was typical for past practices. The lack of a proactive stance, inflexibility and indifference regarding the high-paced ICT development have already been reported as information security problems [18]. It is in fact very likely that, due to some cognitive biases, the management sometimes consciously decides to not put much effort in information security. Additional explanation could be that persons responsible for information security lack the necessary skills. Although decisions regarding investments are taken by company owners, it is still the management’s responsibility to explain why information security is important and, on that basis, establish leadership support for security development. In our opinion, the lack of skilled security practitioners seems to be the most plausible argument for many security-related problems, since this has already proven to be a recognised phenomenon.

Performance and quality measurement is also one of the key factors for developing information security abilities in the right direction and by employing an efficient approach. Information obtained through information security assessment contributes to the reduction of the cognitive bias of decision-makers, user resistance and traditional views. On the contrary, the lack of information about the state-of-play leads to a vicious circle encompassing various problems related to a misaligned ISM (lack of knowledge can lead to overconfidence, budget cuts and incompetence). Despite these arguments, security performance assessments are practically non-existent in many organisations [50], which was also observed in our project.

At this point, other possible perspectives and explanations must also be emphasised. The observed discrepancy between the experts’ recommendations and organisations’ reports in our study is not an absolute indication that the practice is taking a wrong approach. Although our model thrives on well-established guidelines, it is possible that the recommended structure simply does not fit some of the analysed organisations. Since the ISP measurement is a relatively new field of expertise and given the fact that security solutions change constantly, it is actually logical for opposition to arise. A strong indicator supporting such an alternative explanation may be observed in the fact that consensus between experts on the significance of individual processes and controls, or between standards for that matter, has not yet been achieved.

This is why it is difficult to conclude which measures are the most effective. Security assessments, which are currently highly demanded by industry, are somewhat risky, since all effects of individual security controls cannot be measured accurately. When giving recommendations to organisations, we need to be careful and, most importantly, flexible by allowing open interpretations, such as the fact that perceptions can be different and not exclusively right or wrong. The conclusions given in this paper are indeed limited to some extent. Even though we approached the development of the model as objectively as possible and pooled different sources, the recommendations are still limited by certain subjective circumstances, which is in fact the main limitation of the project. Therefore, our conclusions cannot be absolute and we allow the possibility that in a specific situation, the priorities should be set differently.

Nevertheless, the comparison presented in this paper is of great value for both theory and practice, as it explains the possible causes for the ISP deficiencies and unmet security needs in Slovene practice. Although the research studies were conducted in Slovenia, the results have a broader implication, since information security threats and solutions are independent of the environment. This project is one of the steps towards a better understanding of practical dilemmas in ISM. Hopefully, it will facilitate the process of much needed convergence between professionals and practitioners. Particularly the latter should be more included and taken into consideration when developing professional guidelines.

Notes

[1] ISO/IEC 27k, NIST SP 800+, COBIT 5, Critical Security Controls, IASME, PAS55:2013.

[2] Experts evaluated 110 indicators altogether (10 CSFs and 100 KPIs). Factors were graded separately merely for the purpose of calculating weights for KPIs. Further on, in the second research study, when weights were determined, organisations only reported the situation regarding KPIs maturity/development.

[3] Mean values were calculated differently in each research study; in the first study, the average was deducted according to the grade that experts attributed directly to the factor, while mean values from the second research study represented a combination of all ten indicators included in each factor.

[4] The difference between the average value and the value of the gap stems from the fact that the mean represents the non-weighted result, while the gap was calculated on the basis of weighted maximum and average scores. Due to the weighted indicators, every factor has its own maximum limit and every indicator contributes an unequal share. This means that although an entity presents a low overall mean, it could still achieve a high weighted score by developing the most significant indicators within the factor.

[5] By comparison, when analysing differences between average means of all KPIs within a CSF, the highest statistically significant differences were found within the fifth and the eighth factor.

References

[1] R. Baskerville, P. Spagnoletti, and J. Kim, "Information & Management Incident-centered information security : Managing a strategic balance between prevention and response," *Inf. Manag.*, vol. 51, no. 1, pp. 138–151, 2014.

[2] H. Elkhannoubi and M. Belasddaou, "A framework for an effective cybersecurity strategy implementation," *J. Inf. Assur. Secur.*, vol. 11, no. 4, pp. 233–241, 2016.

[3] S. Fenz, J. Heurix, T. Neubauer, and F. Pechstein, "Current challenges in information security risk management," *Inf. Manag. Comput. Secur.*, vol. 22, no. 5, pp. 410–430, Nov. 2014.

[4] Z. A. Soomro, M. H. Shah, and J. Ahmed, “Information security management needs more holistic approach: A literature review,” *Int. J. Inf. Manage.*, vol. 36, no. 2, pp. 215–225, 2016.

[5] M. Zammani and R. Razali, “An Empirical Study of Information Security Management Success Factors,” *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 6, no. 6, p. 904, Dec. 2016.

[6] Gartner, “Gartner Survey Shows Information Security Governance Practices Are Maturing,” 2015. [Online]. Available: http://www.gartner.com/newsroom/id/3098118.

[7] C.-M. Lee and H. Chang, “A study on security strategy in ICT convergence environment,” *J. Supercomput.*, vol. 70, no. 1, pp. 211–223, Oct. 2014.

[8] RSA, “Threat Detection Effectiveness Global Benchmarks 2016,” 2016. [Online]. Available: https://www.rsa.com/content/dam/rsa/PDF/2016/12/h14916-threat-detection-effectiveness-ebook.pdf.

[9] Ponemon Institute and IBM, “The Second Annual Study on the Cyber Resilient Organisation,” 2016. [Online]. Available: http://info.resilientsystems.com/hubfs/IBM_Resilient_Branded_Content/White_Papers/2016_Cyber_Resilient_Organization_Executive_Summary_FINAL.pdf.

[10] Ernst&Young, “Creating trust in the digital world EY’s Global Information Security Survey 2015,” 2015. [Online]. Available: http://www.ey.com/Publication/vwLUAssets/ey-global-information-security-survey-2015/$FILE/ey-global-information-security-survey-2015.pdf.

[11] ISACA, “2015 Global Cybersecurity Status Report,” 2015. [Online]. Available: www.isaca.org/cybersecurityreport.

[12] H. S. B. Herath and T. C. Herath, “IT security auditing: A performance evaluation decision model,” *Decis. Support Syst.*, vol. 57, pp. 54–63, Jan. 2014.

[13] A. Stewart, “Can spending on information security be justified?,” *Inf. Manag. Comput. Secur.*, vol. 20, no. 4, pp. 312–326, Oct. 2012.

[14] M. Burdon, J. Siganto, and L. Coles-Kemp, “The regulatory challenges of Australian information security practice,” *Comput. Law Secur. Rev.*, vol. 32, no. 4, pp. 623–633, Aug. 2016.

[15] J. Jacobs, “Rating the Security Performance of the Fortune 1000,” 2017. [Online]. Available: https://info.bitsighttech.com/rating-the-security-performance-of-the-fortune-1000. [Accessed: 21-Jul-2017].

[16] CESG and CERT-UK, “Common Cyber Attacks: Reducing the Impact,” 2015. [Online]. Available: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/400106/Common_Cyber_Attacks-Reducing_The_Impact.pdf.

[17] Symantec, “Information Security Threat Report. Volume 22.,” 2017. [Online]. Available: https://resource.elq.symantec.com/LP=3980?cid=70138000001BjppAAC&mc=202671&ot=wp&tt=sw&inid=symc_threat-report_regular_to_leadgen_form_LP-3980_ISTR22-report-main.

[18] Verizon, “2017 Data Breach Investigation Report,” 2017. [Online]. Available: http://www.verizonenterprise.com/verizon-insights-lab/dbir/2017/.

[19] FSB, “Cyber Resilience: How to protect small firms in the digital economy,” 2016. [Online]. Available: http://www.fsb.org.uk/docs/default-source/fsb-org-uk/FSB-Cyber-Resilience-report-2016.pdf?sfvrsn=0.

[20] Hiscox, “The Hiscox Cyber Readiness Report 2017,” 2017. [Online]. Available: https://www.hiscox.co.uk/cyber-readiness-report/docs/cyber-readiness-report-2017.pdf.

[21] R. Klahr, J. N. Shah, P. Sheriffs, T. Rossington, and G. Pestell, “Cyber security breaches survey 2017 Main report,” 2017. [Online]. Available: http://www.ipsos-mori.com/terms. [Accessed: 21-Jul-2017].

[22] Centrify and Ponemon Institute, “Ponemon Data Breach Brand Impact UK,” 2017. [Online]. Available: https://www.centrify.com/lp/ponemon-data-breach-brand-impact-uk/.

[23] M. Bromiley, "A SANS Survey Incident Response Capabilities in 2016: The 2016 SANS Incident Response Survey," 2016. [Online]. Available: https://www.sans.org/reading-room/whitepapers/incident/incident-response-capabilities-2016-2016-incident-response-survey-37047.

[24] Venafi, "2016 CIO Study Results. The Threat to Our Cybersecurity Foundation," 2016. [Online]. Available: https://www.venafi.com/assets/pdf/wp/Venafi_2016CIO_SurveyReport.pdf.

[25] ISACA, "State of Cyber Security 2017," 2017. [Online]. Available: https://cybersecurity.isaca.org/state-of-cybersecurity.

[26] PwC, "Global state of information security survey 2017," 2016. [Online]. Available: www.pwc.com/gsiss.

[27] Symantec, "Internet security threat report," 2016. [Online]. Available: https://www.symantec.com/content/dam/symantec/docs/reports/istr-21-2016-en.pdf.

[28] N. K. Sangani and B. Vijayakumar, "Cyber Security Scenarios and Control for Small and Medium Enterprises," *Inform. Econ.*, vol. 16, no. 2, pp. 58–71, 2012.

[29] D. Emm, "Security for SMBs: why it's not just big businesses that should be concerned," *Comput. Fraud Secur.*, vol. 2013, no. 4, pp. 5–8, Apr. 2013.

[30] E. Rohn, G. Sabari, and G. Leshem, "Explaining small business InfoSec posture using social theories," *Inf. Comput. Secur.*, vol. 24, no. 5, pp. 534–556, Nov. 2016.

[31] McAfee, "McAfee Finds Majority of Small Business Owners Have False Sense of Security | McAfee Press Release," *Press Release*, 2013. [Online]. Available: https://www.mcafee.com/us/about/news/2013/q4/20131030-01.aspx.

[32] F. O. Sveen, J. M. Torres, and J. M. Sarriegi, "Blind information security strategy," *Int. J. Crit. Infrastruct. Prot.*, vol. 2, no. 3, pp. 95–109, Oct. 2009.

[33] A. Ahmad, S. B. Maynard, and S. Park, "Information security strategies: towards an organizational multi-strategy perspective," *J. Intell. Manuf.*, vol. 25, no. 2, pp. 357–370, Apr. 2014.

[34] J. P. Pironti, "Developing Metrics for Effective Security Governance | Information Security | Business Process," *Inf. Syst. Control J.*, vol. 2, pp. 1–5, 2007.

[35] B. Barafort, A.-L. Mesquida, and A. Mas, "Integrating risk management in IT settings from ISO standards and management systems perspectives," *Comput. Stand. Interfaces*, vol. 54, no. P3, pp. 176–185, Nov. 2017.

[36] E. Humphreys, "Information security management standards: Compliance, governance and risk management," *Inf. Secur. Tech. Rep.*, vol. 13, no. 4, pp. 247–255, Nov. 2008.

[37] Deloitte, "Central Asian Informations Security Survey Results," 2014. [Online]. Available: https://www2.deloitte.com/content/dam/Deloitte/kz/Documents/risk/KZ_Deloitte_Information_Security_Survey_2014_EN.pdf.

[38] SANS Institute, "IT Security Spending Trends," 2016. [Online]. Available: https://www.sans.org/reading-room/whitepapers/analyst/security-spending-trends-36697.

[39] PwC, "Information security breaches survey 2017," 2017. [Online]. Available: https://www.pwc.be/en/documents/20170315-Information-security-breaches-survey.pdf.

[40] A. Gillies, "Improving the quality of information security management systems with ISO27000," *TQM J.*, vol. 23, no. 4, pp. 367–376, Jun. 2011.

[41] R. Hibbert, "SMBs and the struggle for compliance," *Comput. Fraud Secur.*, vol. 2012, no. 11, pp. 5–7, Nov. 2012.

[42] PwC, "20th CEO Survey - PwC Global," 2017. [Online]. Available: http://www.pwc.com/gx/en/ceo-agenda/ceosurvey/2017/gx.html.

[43] Deloitte and NASCIO, "2014 Deloitte-NASCIO cybersecurity study," 2014. [Online]. Available: https://www2.deloitte.com/us/en/pages/public-sector/articles/2014-deloitte-nascio-cybersecurity-study.html.

[44] M. Siponen and R. Willison, "Information security management standards: Problems and

solutions," *Inf. Manag.*, vol. 46, no. 5, pp. 267–270, Jun. 2009.

[45] D. Olifer, N. Goranin, A. Kaceniauskas, and A. Cenys, "Controls-based approach for evaluation of information security standards implementation costs," *Technol. Econ. Dev. Econ.*, vol. 23, no. 1, pp. 196–219, Jan. 2017.

[46] K. Haufe, R. Colomo-Palacios, S. Dzombeta, K. Brandis, and V. Stantchev, "Security Management Standards: A Mapping," *Procedia Comput. Sci.*, vol. 100, pp. 755–761, 2016.

[47] ISACA, "An Introduction to the Business Model for Information Security," 2009. [Online]. Available: http://www.isaca.org/knowledge-center/research/documents/introduction-to-the-business-model-for-information-security_res_eng_0109.pdf.

[48] A. Guarino, "Information Security Standards in Critical Infrastructure Protection," in *ISSE 2015*, Wiesbaden: Springer Fachmedien Wiesbaden, 2015, pp. 263–269.

[49] I. Bernik and K. Prislan, "Measuring Information Security Performance with 10 by 10 Model for Holistic State Evaluation," *PLoS One*, vol. 11, no. 9, p. e0163050, Sep. 2016.

[50] Ernst&Young, "Insights on governance, risk and compliance. Get ahead of cybercrime EY's Global Information Security Survey 2014," 2014. [Online]. Available: http://www.ey.com/Publication/vwLUAssets/EY-global-information-security-survey-2014/$FILE/EY-global-information-security-survey-2014.pdf.

University of Maribor Press

Explaining the Employment of Information Security Measures by Individuals in Organizations: The Self-protection Model

ANŽE MIHELIČ & SIMON VRHOVEC

Abstract For ensuring information security of an organization, it is important that all its members accessing the cyberspace recognize and deal with cyberthreats individually. Even though appropriate technical cybersecurity measures are commonly employed in organizations, its members are still directly exposed to cyberthreats and are one of the key weak points of ensuring information security of the whole organization. Omission of adequate cybersecurity measures by individuals can be at least embarrassing for the organization and can even critically affect its business in some cases, e.g., through loss of trust. In this paper, we propose a new model based on the insights from extant research in various fields by combining protection motivation theory, technology threat avoidance theory, and psychological reactance theory into a unified model. The unified model explains how different elements influence the intention of individuals to employ information security measures. Its understanding may help information security professionals in adapting information security training to fit the needs of the members of their organizations. Such training that considers the decision-making mechanisms of individuals may stimulate better adoption of information security measures among organization members.

Keywords: • Information security • self-protection • organizations • protection motivation • threat detection • coping •

CORRESPONDENCE ADDRESS: Anže Mihelič, University of Maribor, Faculty of Criminal Justice and Security, Kotnikova 8, 1000 Ljubljana, Slovenia, e-mail: amihelic@student.uni-lj.si. Simon Vrhovec, Ph.D., Assistant Professor, University of Maribor, Faculty of Criminal Justice and Security, Kotnikova 8, 1000 Ljubljana, Slovenia, e-mail: simon.vrhovec@fvv.uni-mb.si.

DOI https://doi.org/10.18690/978-961-286-114-8.2 ISBN 978-961-286-114-8

Available at: http://press.um.si.

1 Introduction

Information-communication technologies (ICT), alongside with the cyberspace, have become an important part of the everyday of any legal entity, whether an individual or an organization. In the case of the latter, the adoption of modern technologies that connect via cyberspace is especially important since the organizations are clamped in electronic commerce at all levels (i.e., B2B, B2G and B2C) as expected by business partners, government, other public authorities and consumers [12].

With the increasing complexity of the technological infrastructure, the difficulty of providing adequate protection against potential cybercrime is growing [46]. The financial loss of companies on a global scale due to information security related issues is already alarming but still increasing [43]. A 2016 study shows average annual losses per single company worldwide now exceed $9.5 million with 21 percent net increase in the total cost over the past year [66].

It is the organization itself that is the first responsible for its information security, by obtaining an adequate technical protection, implementation and adoption of adequate information security policies among its employees. However, it is the portability of modern technologies that enabled people to enter the cyberspace from anywhere that blurred the line between personal and work life [39]. Cybercriminals may employ various techniques of social engineering often in conjunction with technically more or less demanding hacking to attack one of the weakest points in organizations, i.e., their employees. For example, an employee can compromise the organization’s information security just by logging into the organization’s intranet from a home-based computer that is infected with malware. The expansion of the internet access coupled with people’s unawareness of the technology behind the screen ease the intentional exploitation of vulnerabilities and thereby cause a multiplication of negative social and economic consequences [1, 3, 14].

2 Theoretical background

2.1 Protection motivation theory

Protection motivation theory (PMT) was originally developed to help clarify how fear appeals (e.g., warnings) influence ones behavior [51]. In his later revised model, Rogers [52] extended the theory to a more general theory of persuasive communication with an emphasis on the cognitive processes mediating behavioral change [6] in an attempt to explain how individuals change their health attitudes and behaviors in response to health-risk messages.

In general, the revised PMT model suggests two primary cognitive processes that motivate people to engage in actual protection behaviour: threat appraisal and coping appraisal process, and consists of six major components to influence the intention to protect oneself from a threat: (1) perceived severity of the threat; (2) perceived vulnerability (or probability) of the threat; (3) perceived response efficacy of preventive measures; (4) perceived self-efficacy in using preventive measures; (5) rewards; (6) response costs [36, 52]. Due to versatility of the PMT, it has been applied widely to many different fields, especially general health issues [20, 44], food safety [56], safer sex behaviours [2], environmental hazards [61], prevention of nuclear war [4] child protection [15], safe online banking [30], security-related behaviours in organizations [25], information system security [29], self-protection in cyberspace [32, 36] etc.

2.2 Technology threat avoidance theory

Technology threat avoidance theory (TTAT) model explains why and how individuals avoid information technology threats in voluntary settings [38, 39]. This model includes similar constructs with similar names as PMT (e.g., in TTAT, *response efficacy* is renamed to *'safeguard effectiveness'*; *'response cost'* is renamed to *'safeguard cost'*; and *'protection motivation'* becomes *'avoidance motivation'*) also both models share the two primary cognitive processes [16].

The basic premise of TTAT is that when individuals perceive an IT threat, they are motivated to actively avoid the threat by taking a safeguarding measure if they perceive the threat to be avoidable by the safeguarding measure, and they may also passively avoid the threat by performing emotion-focused coping [39]. TTAT includes a process model, a variance model, and many constructs, therefore it is a complex model what is admitted and accurately characterized as complicated by the authors themselves. Complexity of the model can be clearly seen also in other works by other authors as several papers have exhibited a misunderstanding of their model by citing it as a PMT model. TTAT has never been directly compared to the core nomology of PMT and its assumptions [7]. Since there are not many studies related to IT threats, TTAT has expanded the theory by synthesizing various references in the fields of psychology, health care, risk analysis and information system [50].

2.3 Theory of psychological reactance

Psychological reactance [10] is a phenomenon when subject forms a reaction of resistance. The theory suggests that each subject has a freedom of behaviour in any given moment. If any of these behaviours are threatened the subject forms reaction of resistance – reactance [47]. The psychological reactance is an indirect emotional reaction that triggers a resistance to a situation that threatens or limits individual's freedom of behaviour. It appears when someone is forced to accept certain views or enter certain relationships. It is then that it produces a type of behaviour which helps the subject to obtain their freedom. In other words, the subject tries to return his freedom back into the initial state [58]. The subject that has been exposed to stress due to some outside pressure can decide or take the opposite standpoints and demonstrate resistance if the pressure continues [47]. This is often an invisible reaction at first sight. By restraining one's freedom, the individual may repeatedly dishonestly agree to the expectations linked to an authority figure. This may happen in either in a business relationship or in any other non-voluntary interactions which often leads to reduced effectiveness.

A basic condition for forming a reactant response is a contact between two or more subjects. It should be noted that the one causing the stress by limiting the freedom of someone else is not always aware of it. It is however important that the subject feels his freedom is threatened.

The theory of reactance has developed and flourished significantly over the years. Indirectly it is used in health care [21], in campaigns that strive to reduce the usage (consumption) of tobacco products [63], alcohol, drugs and other means to inform the public of a healthy lifestyle [18], marketing or sales strategies [60], social work [54], business and other negotiations [34], sport [11], educational system [24], and also in the field of information security in organizations [40].

3 Self-protection model

In this paper, we propose a novel model of self-protection as presented in Fig. 1. The model is based on the above presented insights and combines the models into a unified theory. It aims to explain how different elements influence the intention of individuals to employ information security measures.

Individuals first need to identify and evaluate information security threats. This is done by assessing their own vulnerability, i.e., the likelihood of information security incidents due to the realization of threats, and the severity of their consequences. The more likely and severe the information security threats the more individuals feel threatened. Therefore, we propose the following hypotheses:

H1a: Perceived vulnerability is positively associated with perceived threat.

H1b: Perceived severity is positively associated with perceived threat.

Fear is one of the more important components of the revised PMT model [41, 53] and its later derivatives [9, 33]. It is commonly used by information security personnel as a tool to achieve the desired behaviour. They try to convince end-users to behave in a more information-secure way through persuasive communication with a certain degree of fear instilling them the fear of loss due to information security incidents. Nevertheless, it is unclear whether such intimidating instructions will even be accepted or how they will be affecting the end-user behaviour [33]. In this paper, we understand fear as an emotional reaction to internet attacks and consequently as a lever for raising the employees' intention to employ information security measures through raising the employees' perceived threat. Hence, we propose the following hypotheses:

H1c: Fear of loss is positively associated with perceived threat.

H2: Perceived threat is positively associated with intention to employ security measures.

Psychological reactance manifests itself when an individual feels that his freedom has been threatened [19, 23, 27]. The more individuals appreciate their freedom the higher the psychological reactance. Similarly, reactance may be higher with higher degree of importance or uniqueness of freedom [55]. Individual traits may also influence psychological reactance, e.g., ethnicity [57, 64], gender [26, 35, 45, 57], age [26, 64] and emotional intelligence [45].

One of the key elements of psychological reactance is perception that an action is not voluntary but mandatory. Mandatoriness could be defined as degree to which individuals perceive that compliance with existing security policies and procedures is compulsory or expected by organizational management [8]. In this paper, we try to distinguish between psychological reactance in general (as a personal trait) and psychological reactance related to information security measures. When evaluating information security measures, individuals assess the degree of freedom they have in adopting them. In case the measures are imposed, individuals may feel their freedom to choose has been tampered and they resist their adoption. Even if the measures are adopted, individuals may not entirely embrace them which is similar to resistant use in change management theory [48, 62]. In such cases, information security measures may be employed however not as efficiently as they could be. We developed the following hypotheses:

> **H3a:** Psychological reactance is positively associated with psychological reactance of security measures.
>
> **H3b:** Mandatoriness of the security measures is positively associated with psychological reactance of security measures.
>
> **H4:** Psychological reactance of security measures is negatively associated with intention to employ security measures.

Finally, drawing on prior research on health protective behavior [31, 42] and self-efficacy [5, 17] as summarized in TTAT [38], individuals consider three factors to evaluate how avoidable the threat can be made by a safeguarding measure: the effectiveness of the measure, the costs of the measure, and individuals' self-efficacy of taking the measure. We assume that they are associated with intention to employ security measures directly, since they are all regarded as key factors affecting the frequency of responding to information security threats [13, 37]. Finally, we develop the last three hypotheses:

> **H5:** Self-efficacy is positively associated with intention to employ security measures.
>
> **H6:** Measure efficacy is positively associated with intention to employ security measures.
>
> **H7:** Costs are positively associated with intention to employ security measures.

The proposed self-protection model is presented in Fig. 1.

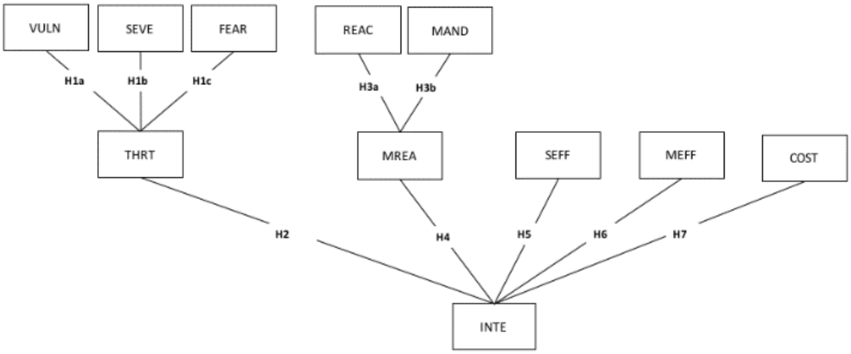


Figure 1: Model for self-protection against information security threats.

4 Methodology

We conducted a survey in six Slovenian universities to test the proposed model. We administered the online questionnaire to 3,872 publicly available e-mail addresses of teaching staff at the following universities: University of Ljubljana, University of Maribor, University of Nova Gorica, University of Primorska, University of Novo mesto and Higher Education Centre Novo mesto. A total of 285 respondents completed the survey providing a response rate of 7.7 percent. 56.1 percent of the respondents were females.

Survey questionnaire was designed to measure eleven constructs: perceived vulnerability (VUL), perceived severity (SEV), fear of loss (FEAR), perceived threat (THRT), potential psychological reactance (REAC), psychological reactance of security measure (MREA), mandatoriness (MAND), measure efficacy (MEFF), self-efficacy (SEFF), expected measure costs (COST), and intention to employ the measure (INTE). The used survey items were all previously validated items adapted to relate specifically to the context of our model. VUL and SEV items were adapted from Liang and Xue [39], FEAR items were adapted from Osman et al. [49], CONC items were adapted from Herath and Rao [25], REAC and MREA items were adapted from Hong and Page [28], MAND items were adapted from Boss et al. [8], MEFF items were adapted from Chen and Zahedi [16], SEFF and INTE items were adapted from Taylor and Todd [59], and COST items were adapted from Woon, Tan and Low [65]. All items were measured using a seven-point Likert scale from I strongly disagree (1) to I strongly agree (7).

The reliability analysis showed that all constructs had adequate reliability. Most constructs had Cronbach's alpha values above 0.8, except for REAC (0.727) and SEFF (0.741), indicating a high degree of reliability. To test the research hypotheses, we analyzed the correlations between constructs by calculating Spearman's rho which is a nonparametric method for analyzing data appropriate when the items are not normally distributed.

5 Results

To test the hypotheses, we calculated Spearman's rho rank monotonic correlation coefficients as REAC, MREA and MEFF were not following a normal distribution. Figure 2 presents the proposed model with the results of hypotheses testing.

The first set of hypotheses tries to explain employee's perceived threat and its relation to intention to employ information security measures. In hypotheses H1a, H1b and H1c, we expected positive association between constructs VULN and THRT, SEVE and THRT, and FEAR and THRT, respectively. The results showed that all correlations are positive and statistically significant ($p < .001$). Based on these results, we can confirm hypotheses H1a, H1b and H1c. In hypothesis H2, we anticipated that THRT is positively associated with INTE. Analysis showed that there is a statistically significant positive correlation between the constructs ($p < .001$). Therefore, we can also confirm hypothesis H2.

The second set of hypotheses tries to explain the role of psychological reactance and mandatoriness of taking the information security measures. In hypotheses H3a and H3b, we anticipated positive associations between REAC and MREA, and between MAND and MREA. Analysis of the results showed positive correlation between REAC and MREA ($p < .001$), and statistically significant negative correlation between MAND and MREA ($p < .05$). We can confirm hypothesis H3a based on these results and reject hypothesis H3b as the correlation is negative. In hypothesis H4, we assumed a negative association between MREA and INTE. Analysis showed a negative correlation coefficient however it is not statistically significant ($p = .098$). Based on these results, we cannot confirm or reject hypothesis H4. We additionally calculated the correlation between MAND and INTE ($r_s = .242$, $p < .05$).

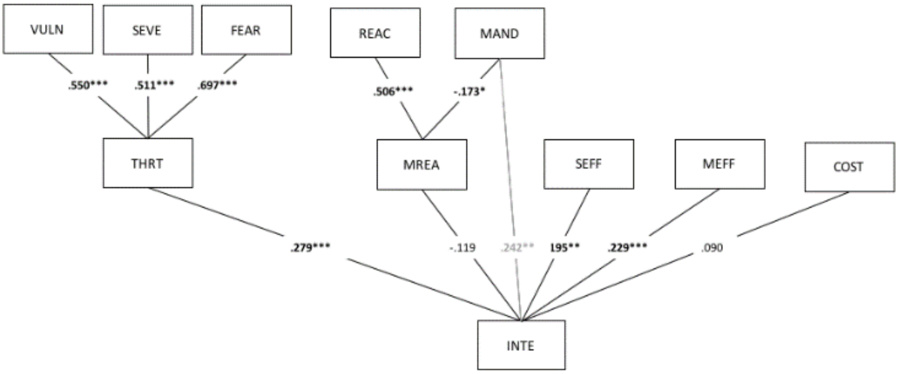


Figure 2: Hypotheses testing results Spearman's rho rank correlation coefficients (*: p < .05, **: p < .01, ***: p < .001).

The third set of hypotheses tries to explain how measure attributes relate to intention to employ information security measures. In hypotheses H5, H6 and H7, we expected a positive correlation between SEFF and INTE, MEFF and INTE, and COST and INTE. Analysis showed a statistically significant correlation between SEFF and INTE ($p < .01$), and MEFF and INTE ($p < .001$) therefore confirming hypotheses H5 and H6. The correlation between COST and INTE was however not significant. Thus, we cannot neither confirm nor reject hypothesis H7.

6 Discussion

The findings of this paper are confirming most assumptions of the proposed model. Nevertheless, one hypothesis was rejected and two hypotheses could not be neither confirmed nor rejected. In hypothesis H3b, we predicted that MAND is positively associated with MREA. We postulated that if someone perceives a high degree of mandatoriness, i.e., lower freedom of choosing whether to employ a security measure or not, then his reactance of these security measures would be higher too. The results however showed quite the opposite as there is a significant negative correlation between the two. These results are not easily explained. It could be due to a non-linear reactance curve. Reactance is increasing when freedom is being limited up to a certain point and then starts decreasing. This would however mean that universities are some kind of a total institution as suggested by some authors [22] with a low degree of freedom when considering employing information security measures. It would be interesting to see if research in different settings would also yield the same results.

We also failed to confirm hypothesis H4 that predicted a negative association between MREA and INTE. This failure could also be attributed to the non-linear reactance curve as described above. Alternatively, the theory on psychological reactance may not be applicable in these settings. Further work on both general and security measure reactance curves would be needed to determine the causes. Repeating the research in different settings would also be beneficial. There is however a positive correlation between MAND and INTE. This is an approximate answer to probably one of the questions that managers are very interested in, i.e., does the mandatoriness of an information security measure correlate with the intention to employ it. Further work may explore the viability to turn this part of the model around. Such a model would try to relate reactance to intention to employ a security measure through its perceived mandatoriness.

The failure to confirm hypothesis H7 seems contrary to well established theories [36, 38]. This could be attributed to people not perceiving the investment of time, effort and workload into protecting their devices as considerable costs. Additional research would be needed to determine whether this is the case or the costs are mediated by perceived benefits of employing security measures.

The reader should note that the study has been done in university settings. Different settings might yield different results therefore further research in various settings would be beneficial.

7 Conclusion

The proposed model deals with self-protecting behavior of individuals. It empowers organizations to better adapt their information security policies and training to fit the needs of their employees and their self-protecting behavior. This improves the adoption of information security measures by employees and thus the overall information security of the organization. Individuals may be more motivated to employ information security measures not only at work but also in their personal life.

References

[1] Acquisti, A., Friedman, A. and Telang, R. 2006. Is There a Cost to Privacy Breaches? An Events Study. *Fifth Workshop on the Economics of Information Security* (2006), 1--20.

[2] Ahia, R.N. 1991. Compliance with safer-sex guidelines among adolescent males: application of the health belief model and protection motivation theory. *Journal of Health Education*. 22, (1991), 49–52.

[3] Anderson, C.L. and Agarwal, R. 2010. Practicing Safe Computing: A Multimethod Empirical Examination of Home Computer User Security Behavioral Intentions. *MIS Quarterly*. 34, 3 (2010), 613–643.

[4] Axelrod, J.J. and Newton, J.W. 1991. Preventing nuclear war: beliefs and attitudes as predictors of disarmist and deterrentist behaviour. *Journal of Applied Social Psychology*. 133, (1991), 459–467.

[5] Bandura, A. 1982. Self-Efficacy Mechanism in Human Agency. *American Psychologist*. 37, (1982), 122–147.

[6] Boer, H. and Seydel, E.R. 1996. Protection motivation theory. *Predicting Health Behavior*. M. Connor and P. Norman, eds. Open University Press. 95–120.

[7] Boss, S.R., Galletta, D.F., Lowry, P.B., Moody, G.D. and Polak, P. 2015. What do systems users have to fear? Using fear appeals to engender threaths and fear that motivate protective security behaviors. *MIS Quarterly*. 39, 4 (2015), 837–864.

[8] Boss, S.R., Kirsch, L.J., Angermeier, I., Shingler, R.A. and Boss, R.W. 2009. If someone is watching, I'll do what I'm asked: mandatoriness, control, and information security. *European Journal of Information Systems*. 18, (2009), 151–164. DOI:https://doi.org/10.1057/ejis.2009.8.

[9] Boss, S.R., Street, F., Galletta, D.F., Lowry, P.B., Kong, H., Moody, G.D. and Polak, P. 2015. What do systems users have to fear? Using fear appeals to engender threaths and fear that motivate protective security behaviors. *MIS Quarterly*. 39, 4 (2015), 1–13.

[10] Brehm, S.S. and Brehm, J.W. 1981. *Psychological Reactance: A Theory of Freedon and Control*. Academic Press.

[11] Briki, W., Den Harigh, R., Hauw, D. and Gernigon, C. 2012. A qualitative exploration of the psychological contents and dynamics of momentum in sport. *International Journal of Sport Psychology*. 43, 5 (2012), 365–384. DOI:https://doi.org/10.7352/IJSP2012.43.365.

[12] Browne, S., Lang, M. and Golden, W. 2015. Linking Threat Avoidance and Security Adoption : A Theoretical Model For SMEs. *#eWellBeing* (2015), 32–43.

[13] Browne, S., Lang, M. and Golden, W. 2015. Linking Threat Avoidance and Security Adoption : A Theoretical Model For SMEs. *#eWellBeing* (2015), 32–43.

[14] Campbell, K., Gordon, L.A., Loeb, M.P. and Zhou, L. 2003. The economic cost of publicly announced information security breaches: empirical evidence from the stock market. *Journal of Computer Security*. 11, 3 (2003), 431–448.

[15] Campis, L.K., Prentice-Dunn, S. and Lyman, R.D. 1989. Coping appraisal and parents' intention to inform their children about sexual abuse: a protection motivation theory analysis. *Journal of Social and Clinical Psychology*. 8, (1989), 304–316.

[16] Chen, Y. and Zahedi, F.M. 2016. Individuals' internet security perceptions and behaviors: Polycontextual contrasts between the United States and China. *MIS Quarterly*. 40, 1 (2016), 205–222.

[17] Compeau, D.R. and Higgins, C.A. 1995. Computer Self- Efficacy: Development of a Measure and Initial Test. *MIS Quarterly*. 19, (1995), 189–211.

[18] Compton, J., Jackson, B. and Dimmock, J.A. 2016. Persuading Others to Avoid Persuasion: Inoculation Theory and Resistant Health Attitudes. *Frontiers in Psychology*. 7, February (2016), 1–9. DOI:https://doi.org/10.3389/fpsyg.2016.00122.

[19] Dowd, E.T., Milne, C.R. and Wise, S.L. 1991. The therapeutic reactance scale: A measure of psychological reactance. *Journal of Counseling & Development.*

[20] Flynn, M.F., Lyman, R.D. and Prentice-Dunn, S. 1995. Protection motivation theory and adherence to medical treatment regimens for muscular dystrophy. *Journal of Social and Clinical Psychology*. 14, (1995), 61–75.

[21] Fogarty, J.S. 1997. Reactance theory and patient noncompliance. *Social Science and Medicine*. 45, 8 (1997), 1277–1288. DOI:https://doi.org/10.1016/S0277-9536(97)00055-5.

[22] Gibbon, H.M.F., Canterbury, R.M. and Litten, L. 1999. Colleges as total institutions: implications for admission, orientation, and student life. *College and University*. 74, 2 (1999), 21–27.

[23] Goldsmith, R.E. and Clark, R.A. 2005. Tendency to conform: A new measure and its relationship to psychological reactance. *Psychological Reports*. 96 (2005), 591–594.

[24] Graca, J., Calheiros, M.M. and Barata, M.C. 2013. Authority in the classroom: Adolescent autonomy, autonomy support, and teachers' legitimacy. *European Journal of Psychology of Education*. 28, 3 (2013), 1065–1076. DOI:https://doi.org/10.1007/s10212-012-0154-1.

[25] Herath, T. and Rao, H.R. 2009. Protection Motivation and Deterrence: a Framework for Security Policy Compliance in Organisations. *European Journal of Information Systems*. 18, 2 (2009), 106–125. DOI:https://doi.org/10.1057/ejis.2009.6.

[26] Hong, S.M., Giannakopoulos, E., Laing, D. and Williams, N. a 1994. Psychological reactance: effects of age and gender. *The Journal of social psychology*. 134, 2 (1994), 223–228. DOI:https://doi.org/10.1080/00224545.1994.9711385.

[27] Hong, S.M. and Page, S. 1989. A Psychological Reactance Scale - Development, Factor Structure and Reliability. *Psychological Reports*. 64, 3 (1989), 1323–1326.

[28] Hong, S.M. and Page, S. 1989. A Psychological Reactance Scale - Development, Factor Structure and Reliability. *Psychological Reports*. 64, 3 (1989), 1323–1326.

[29] Ifinedo, P. 2014. Information systems security policy compliance: An empirical study of the effects of socialisation, influence, and cognition. *Information and Management*. 51, 1 (2014), 69–79. DOI:https://doi.org/10.1016/j.im.2013.10.001.

[30] Jansen, J. 2015. Studying Safe Online Banking Behaviour: A Protection Motivation Theory Approach. *Proceedings of the Ninth International Symposium on Human Aspects of Information Security & Assurance (HAISA 2015) Studying*. Haisa (2015), 120–130.

[31] Janz, N.K. and Becker, M.H. 1984. The Health Belief Model: A Decade Later. *Health Education Quarterly*. 11, 1 (1984), 1–45.

[32] Jenkins, J.L., Grimes, M., Proudfoot, J.G. and Lowry, P.B. 2013. Improving Password Cybersecurity Through Inexpensive and Minimally Invasive Means: Detecting and Deterring Password Reuse Through Keystroke-Dynamics Monitoring and Just-in-Time Fear Appeals. *Information Technology for Development*. 20, 2 (2013), 196–213. DOI:https://doi.org/10.1080/02681102.2013.814040.

[33] Johnston, B.A.C. and Warkentin, M. 2010. Fear appeals and information security

behaviors: An empirical study. *MIS Quarterly*. 34, 3 (2010), 549–566.

[34] Kray, J.L., Thompson, L. and Galinsky, A. 2001. Battle of the sexes: gender stereotype confirmation and reactance in negotiations. *Journal of Personality and Social Psychology*. 80, 6 (2001), 942–958.

[35] Kray, J.L., Thompson, L. and Galinsky, A. 2001. Battle of the sexes: gender stereotype confirmation and reactance in negotiations. *Journal of Personality and Social Psychology*. 80, 6 (2001), 942–958.

[36] Lee, D., Larose, R. and Rifon, N. 2008. Keeping our network safe: a model of online protection behaviour. *Behaviour & Information Technology*. 27, 5 (2008), 445–454. DOI:https://doi.org/10.1080/01449290600879344.

[37] Lee, D., Larose, R. and Rifon, N. 2008. Keeping our network safe: a model of online protection behaviour. *Behaviour & Information Technology*. 27, 5 (2008), 445–454. DOI:https://doi.org/10.1080/01449290600879344.

[38] Liang, H. and Xue, Y. 2009. Avoidance of Information Technology Threats: A Theoretical Perspective. *MIS Quarterly*. 33, 1 (2009), 71–90. DOI:https://doi.org/Article.

[39] Liang, H. and Xue, Y. 2010. Understanding Security Behaviors in Personal Computer Usage: A Threat Avoidance Perspective. *Journal of the Association for Information Systems*. 11, 7 (2010), 394–413.

[40] Lowry, P.B. and Moody, G.D. 2014. Proposing the control-reactance compliance model (CRCM) to explain opposing motivations to comply with organisational information security policies. *Information Systems Journal*. May (2014), 433–463. DOI:https://doi.org/10.1111/isj.12043.

[41] Maddux, J.E. and Rogers, R.W. 1983. Protection motivation theory and self-efficacy: A revised theory of fear appeals and attitude change. *Journal of Experimental Social Psychology*. 19, (1983), 469–479.

[42] Maddux, J.E. and Rogers, R.W. 1983. Protection motivation theory and self-efficacy: A revised theory of fear appeals and attitude change. *Journal of Experimental Social Psychology*. 19, (1983), 469–479.

[43] McAffe 2013. *The Economic impact of cyber crime and cyber espionage*.

[44] Mesters, I., Meertens, R., Kok, G. and Percel, G.S. 1994. Effectiveness of a multidisciplinary education protocol on children with asthma (0–4 years) in primary health care. *Journal of Asthma*. 31, (1994), 347–359.

[45] Middleton, J., Buboltz, W. and Sopon, B. 2014. The relationship between psychological reactance and emotional intelligence. *Social Science Journal*. 52, 4 (2014), 542–549. DOI:https://doi.org/10.1016/j.soscij.2015.08.002.

[46] Mihelič, A. and Vrhovec, S. 2016. *Soočanje z najpogostejšimi ranljivostmi spletnih aplikacij državnih organov*.

[47] Miron, A.M. and Brehm, J.W. 2006. Reactance Theory - 40 Years Later. *Zeitschrift für Sozialpsychologie*. 37, 1 (2006), 9–18. DOI:https://doi.org/10.1024/0044-3514.37.1.9.

[48] van Offenbeek, M., Boonstra, A. and Seo, D. 2013. Towards integrating acceptance and resistance research: evidence from a telecare case study. *European Journal of Information Systems*. 22, 4 (2013), 434–454. DOI:https://doi.org/10.1057/ejis.2012.29.

[49] Osman, A., Barrious, F.X., Osman, J.R., Schneekloth, R. and Troutman, J.A. 1994. The Pain Anxiety Symptoms Scale: Psychometric Properties in a Community Sample. *Journal of Behavioral Medicine*. 17, 5 (1994), 511–522.

[50] Rho, H. and Yu, I. 2011. The impact of information technology threat avoidance factors on avoidance behavior of user. *Department of Business Management, SunCheon National University*. (2011).

[51] Rogers, R.W. 1975. A protection motivation theory of fear appeals and attitude change. *Journal of Psychology*. 91, (1975), 93–114.

[52] Rogers, R.W. 1983. Cognitive and physiological process in fear appeals and attitude

change: a revised theory of protection motivation. *Social Psychophysiology: a source book*. J. Cacioppo and R. Petty, eds. Guilford Press. 153–176.

[53] Rogers, R.W. 1983. Cognitive and physiological process in fear appeals and attitude change: a revised theory of protection motivation. *Social Psychophysiology: a source book*. J. Cacioppo and R. Petty, eds. Guilford Press. 153–176.

[54] Rooney, H.R. 2009. *Strategies for Work with Involuntary Clients*. Columbia University Press.

[55] Rooney, H.R. 2009. *Strategies for Work with Involuntary Clients*. Columbia University Press.

[56] Schafer, R.B., Schafer, E., Bultena, G. and Hoiberg, E. 1993. Coping with health threat: a study of food safety. *Journal of Applied Social Psychology*. 23, (1993), 386–394.

[57] Seemann, E. a., Buboltz, W.C., Jenkins, S.M., Soper, B. and Woller, K. 2004. Ethnic and gender differences in psychological reactance: the importance of reactance in multicultural counselling. *Counselling Psychology Quarterly*. 17, 2 (2004), 167–176. DOI:https://doi.org/10.1080/09515070410001728316.

[58] Šugman Bohic, L. 2008. Ocenjevanje začetnih stikov v neprostovoljnih transakcijah.

[59] Taylor, S. and Todd, P.A. 1995. Understanding Information Technology Usage - a Test of Competing Models. *Information Systems Research*. 6, 2 (1995), 144–176.

[60] Trump, R.K. 2016. Harm in price promotions: when coupons elicit reactance. *Journal of Consumer Marketing*. (2016). DOI:https://doi.org/10.1108/JCM-02-2015-1319.

[61] Vaughn, E. 1993. Chronic exposure to environmental hazard: risk perception and self-protective behavior. *Health Psychology*. 12, (1993), 74–85.

[62] Vrhovec, S.L.R. 2016. Responding to stakeholders' resistance to change in software projects - A literature review. *39th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO 2016)* (Opatija, Croatia, 2016).

[63] Wiium, N., Aarø, L.E. and Hetland, J. 2009. Psychological reactance and adolescents' Attitudes toward tobacco-control measures. *Journal of Applied Social Psychology*. 39, 7 (2009), 1718–1738. DOI:https://doi.org/10.1111/j.1559-1816.2009.00501.x.

[64] Woller, K.M.P., Buboltz, W.C. and Loveland, J.M. 2007. Psychological reactance: Examination across age, ethnicity, and gender. *The American Journal of Psychology*. 120, 1 (2007), 15–24.

[65] Woon, I., Tan, G.-W. and Low, R. 2005. A Protection Motivation Theory Approach to Home Wireless Security. *Proceedings of the 26th International Conference on Information Systems* (2005), 367–380.

[66] 2016. *2016 Cost of Cyber Crime Study & the Risk of Business Innovation*.

Concept Drift Analysis for Improving Anomaly Detection Systems in Cybersecurity

MICHAŁ CHORAŚ & MICHAŁ WOŹNIAK

Abstract In this paper we propose to add concept drift analysis as the pre-processing step dedicated for anomaly detection systems to counter cyber attacks. Such approach could be a step towards lifelong learning intelligent cyber security system. Such system could use the previously learnt knowledge to properly detect cyber attacks in the ever changing networked environments without the necessity to re-learn from scratch. We believe that such approach could improve the detection rate e.g. of the obfuscated SQL injection attacks and decrease the false positive rates by properly adjusting to the concept drift in the data.

Keywords: • Cyber security • machine learning • data science • concept drift • anomaly detection •

CORRESPONDENCE ADDRESS: Michał Choras, Ph.D., Professor, UTP University of Science and Technology, Institute of Telecommunications, Al. prof. S. Kaliskiego 7, 85-796 Bydgoszcz, Poland, e-mail: chorasm@utp.edu.pl. Michał Woźniak, Ph.D., Professor, Wrocław University of Technology, Faculty of Electronics, Department of Systems and Computer, 11/17 Janiszewskiego st., 50-372 Wrocław, Poland, e-mail: michal.wozniak@pwr.edu.pl.

DOI https://doi.org/10.18690/978-961-286-114-8.3 ISBN 978-961-286-114-8

Available at: http://press.um.si.

1 Introduction

Network and information security is now one of the most prominent problems of citizens, societies and homeland security [1]. As widely observed, the number of successful attacks on information, citizens and even secure financial systems, as well as critical infrastructures is still growing [2][3]. One of the problem lies with the inefficiency of signature based approaches to detect cyber attacks. In situations, where new attacks (or even slightly modified families of malware) emerge, those systems are not efficient until the new signatures are created [4]. On the other hand, anomaly based approaches (systems which detect abnormalities in traffic or e.g. requests to databases) tend to produce false positives (false alarms).

Such situation and current challenges in network security motivate our research and the concept to apply concept drift detectors and the lifelong learning approach to cybersecurity domain.

Of course, we do not suggest to replace, but rather to complement the traditional signature-based and anomaly-based solutions by our approach.

2 Towards applying concept drift in cyber security

2.1 Concept drift

Concept drift is a change in the characteristics of the data stream. It means that the characteristics of the decision attributes and of the classes to be predicted, change in time in unpredictable manner. Such situation may cause the decrease in classification quality. What is interesting, the classification may further decrease with each new portion of the data (in contrast to situation where new data should increase the classification quality).
In cybersecurity context, it would mean the decrease of cyber attacks detection in time, due to the natural (not malicious) and unpredictable changes of network traffic characteristics.

Therefore, we think that the concept drift detection techniques should be applied to the autonomous detection of the model parameter changes related to incoming new traffic (and new learning tasks).

2.2 Concept Drift taxonomy

Concept drift [9] may come in many forms, depending on the type of change. Usually, its appearance spoils quality of used models, therefore, developing such methods which can effectively deal with this phenomenon is still a focus of intense research. There are a few taxonomies of concept drift, but let’s focus on two of them. We can categorize drifts according their influences into probabilistic characteristics of a classification [16], i.e.:

- virtual concept drift does not have any impact on decision boundaries [11].
- real concept drift has an impact on decision boundaries.

The real concept drift is the most important, but detecting the virtual one may be also useful, especially in the case if the drift detection may cause model rebuilding, because in the case of virtual drift (which does change the decision boundaries) such an action is not necessary.

We may also distinguish the drift types according to the drift impetuosity:

- gradual or incremental drift - for the gradual drift (see. Fig.1a) for a given period of time examples from different models may appear in the stream concurrently, while for incremental drift (see. Fig.1b) the model's parameters are changing smoothly.
- sudden drift, where the drift has rapid nature (see Fig.1c) and sometimes the old model may appear again (see. Fig.1d).

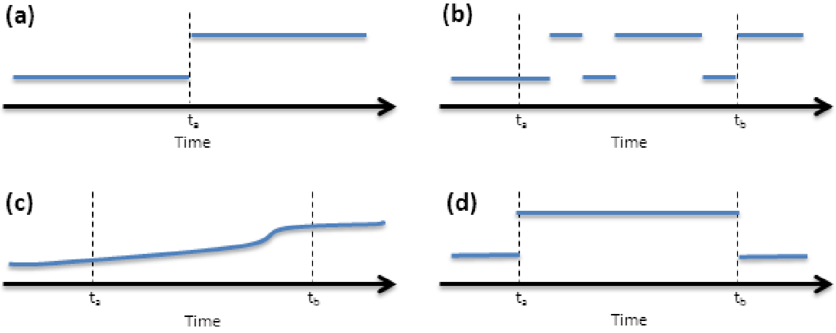


Figure 1 Types of changes in the data stream: (a) sudden, (b) gradual, (c) incremental, (d) recurring

Concept drift detector is an algorithm, which on the basis of information about incoming observation and model's performance can return signal that data stream distributions are changing, usually they can return the signal that drift is detected and model should be rebuilt as quick as possible or that so-called warning level is achieved, what may cause the necessity to collect new data to rebuild/update the model.

The drift detector could be recognized as the simple classifier, but from practical point of view they rather solve the regression problem, evaluation how far the used model is from the model of the real problem. Such a task is tough, because on the one hand we should detect a drift as soon as possible to replace outdated model and to reduce so-called restoration time, but on the other hand we do not accept too many false alarms [17].

Additionally, it is worth noticing that drift detectors are usually assuming the continue access to class labels, which cannot be granted from the practical point of view. Therefore, building such systems we should take into consideration the data labeling cost, which is usually passed over. Unfortunately, without access to class labels the real drift could be undetected [18].

3 Proposal for concept drift in cybersecurity

Cyber attacks detection methods and systems can be as weak, as weak are the advanced data processing approaches and algorithms. The standard approach to cyber attacks detection is so-called signature-based mode, where the patterns of 'evil' malicious traffic are compared to current traffic samples, and if matched, the alarm is raised (Fig. 2). Such approach is, of course, inefficient in detection of the new or modified cyber attacks, or so-called 0-day exploits.

Therefore, another approach is to detect anomalies. To do so, firstly, the pattern of normality (normal traffic etc.) has to be learnt and then matched versus the current traffic samples. Whenever, there is no match, the alarm is raised (Fig. 3). This approach is however plagued by false positives (false alarms). Quite often, when the characteristics of network traffic (or e.g. HTTP requests in the application layer) change, such situation is detected as anomaly, even though it is just a normal change of the network behavior. In pattern recognition, such situations are termed as concept drift, and this aspect should be taken into account in detection systems. Therefore in this paper, we postulate to include concept drift detector as kind of the pre-processing step in anomaly detection cyber security systems (Fig. 4).

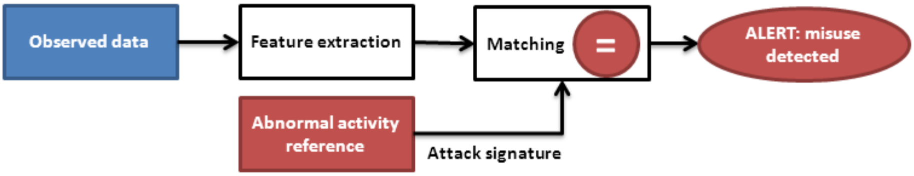


Figure 2 Signature-based approach for cyber attacks detection

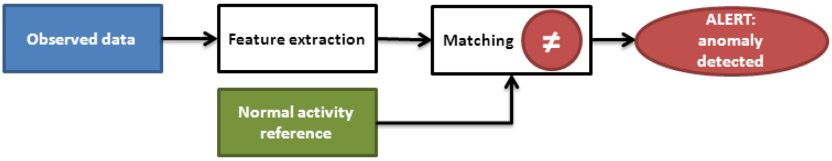


Figure 3 Anomaly detection approach for cyber security detection

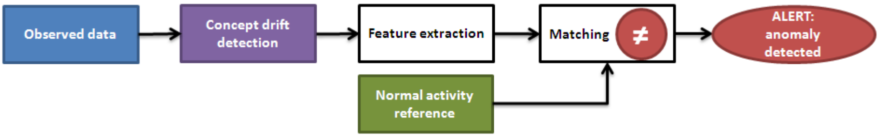


Figure 4 Anomaly detection approach enhanced with the concept drift detectiom

3.1 Towards Lifelong Learning Intelligent Systems in Cybersecurity

The successful application of the Concept Drift Detectors is a step forwards the cybersecurity lifelong learning intelligent system [19].

One of the main drawbacks of many cybersecurity solutions adapting machine learning is the fact that the learning process is concerned as STL (Single Task Learning problem). For instance, when developing an anomaly detection system (ADS), researchers usually [5] collect different malicious network traffic samples generated by different malware families, then split the data for training and evaluation (often inspecting manually the data to ensure that the datasets will be similarly distributed), and finally train the model. The following aspects shall be considered:

- The tasks collecting data samples for different malware families are not identical and thus those should not be treated as single task;
- On the other hand, while treating those tasks separately (learning the classifiers independently) the information that could be acquired from other tasks will be missing.

When it comes to the cybersecurity and cyber-attacks detection, there is no single classifier or IDS system that will allow recognizing all kinds of attacks. Also the same system (even if learnt to detect the same type of attacks) have to be learnt again when changing the monitored network (topology, services, characteristics etc.). In that regards we will need transfer learning mechanism that will allow us to learn to detect attack B from knowledge (e.g. ensembles of classifiers) learnt for attack A.

Another aspect is the overlap of knowledge that our system will need to be aware of. We will leverage this both to facilitate learning of new tasks and improving the effectiveness, when executing the old ones. Using again the cybersecurity example, IDS and/or anomaly detection system learnt in one network will use already established knowledge to detect attacks in another new network in a more accurate way (than without lifelong learning approach).

4 Future work and use cases

We currently focus on adaptation of concept drift detection mechanism in the area of cybersecurity. In particular, we plan to adapt the above-mentioned ideas to efficiently detect anomalies in the monitored networks.

Application layer attacks, such as SQLIA (SQL Injection Attacks) are top-ranked on several threat lists. One of the examples is the "OWASP Top 10" [9] list that has been identified by Open Web Application Security). The list, among others, contains the following items from the application layer: injection flaws (e.g. SQL Injection), broken authentication and session management, Cross-Site Scripting.

The practical implementation of concept drift detection and anomaly detection (as in Fig. 4) approach to protect the application layer can consider for example the analysis of user requests to web service or data-base. In such scenario, we can apply sensors and implement complex algorithms to learn the models of normal requests or user behaviour, and detect all the requests that fall outside the model of normality.

However, in order to decrease the false positive ratio (indicating anomalies which are not attacks or symptoms of misbehaviour) we now work to implement concept drift detectors and further apply lifelong learning solutions.

In such scenario, the model of normal requests changes quickly (e.g. due to availability of new services or just new fields in web forms) and therefore anomaly detection approach tends to have high false positive ratio. In our proposed solution, the system will:

- Detect concept drifts
- React to concept drifts
- And re-learn using the past knowledge to quickly adapt to network changes.

5 Conclusions

In summary, in this paper we have presented the general concept of applying concept drift detection to cybersecurity domain. We have proposed the new general framework of the anomaly detection system enriched by concept drift detection. Such solution is a step towards lifelong learning intelligent system adopted to cybersecurity domain.

We currently work on collecting the relevant data (there are not many relevant benchmark databases yet available) and on the experimental evaluation of the proposed approach.

References

[1] Choraś M., Kozik R., Torres Bruna M.P., Yautsiukhin A., Churchill A., Maciejewska I., Eguinoa I., Jomni A.: Comprehensive Approach to Increase Cyber Security and

Resielience, in Proc. of ARES (International Conference on Availability, Reliability and Security), 686-692, Touluse, IEEE (2015)

[2] Kozik, R., Choraś, M., Renk, R., Hołubowicz, W.: Cyber Security of the Ap-plication Layer of Mission Critical Industrial Systems, in: Saeed K. and Homenda W., Computer Information Systems and Industrial Management, CISIM 2016, Vilnius, Lecture Notes in Computer Science LNCS 9842, 342-351 (2016)

[3] Choraś, M., Kozik, R., Flizikowski, A., Renk, R., Hołubowicz, W.: Cyber Threats Impacting Critical Infrastructures, in: Setola R. et al.: Managing the Complexity of Critical Infrastructures, Studies in Systems, Decision and Con-trol, vol. 90, 139-161 (2017)

[4] Choraś, M., Kozik, R., Puchalski, D., Hołubowicz, W.: Correlation Approach for SQL Injection Attacks Detection, In: Herrero A. et al (Eds.), Advances in Intelligent and Soft Computing, 189, 177-186 (2012)

[5] Sebastián García, Alejandro Zunino, Marcelo Campo: Survey on network-based botnet detection methods. Security and Communication Networks 7(5): 878-903 (2014)

[6] Hoens, T. Ryan, Robi Polikar, and Nitesh V. Chawla. "Learning from streaming data with concept drift and imbalance: an overview." Progress in Artificial Intelligence 1.1 (2012): 89-101.

[7] Tsymbal, Alexey. "The problem of concept drift: definitions and related work." Computer Science Department, Trinity College Dublin 106.2 (2004).

[8] Widmer, Gerhard, and Miroslav Kubat. "Learning in the presence of concept drift and hidden contexts." Machine learning 23.1 (1996): 69-101.

[9] OWASP – The Open Web Application Project – OWASP Top Ten.

[10] Gama, J., Žliobaitė, I., Bifet, A., Pechenizkiy, M., & Bouchachia, A. (2014). A survey on concept drift adaptation. ACM Computing Surveys (CSUR), 46(4), 44.

[11] Gerhard Widmer and Miroslav Kubat. Effective learning in dynamic environments by explicit context tracking. In PavelB. Brazdil, editor, Machine Learning: ECML-93, volume 667 of Lecture Notes in Computer Science, pages 227-243. Springer, Berlin Heidelberg, 1993.

[12] Orij, J. (2016). Self-adaptation to concept drift in web-based anomaly detection (Master's thesis, University of Twente).

[13] Demertzis, K., & Iliadis, L. (2013, December). A hybrid network anomaly and intrusion detection approach based on evolving spiking neural network classification. In International Conference on e-Democracy (pp. 11-23). Springer

[14] Wang, W., Guyet, T., Quiniou, R., Cordier, M. O., Masseglia, F., & Zhang, X. (2014). Autonomic intrusion detection: Adaptively detecting anomalies over unlabeled audit data streams in computer networks. Knowledge-Based Systems, 70, 103-117.

[15] Lampesberger, H., Winter, P., Zeilinger, M., & Hermann, E. (2011, September). An On-Line Learning Statistical Model to Detect Malicious Web Requests. In SecureComm (pp. 19-38).

[16] J. Gama, I. Zliobaite, A. Bifet, M. Pechenizkiy, and A. Bouchachia. A survey on concept drift adaptation. ACM Computing Surveys, volume 46 Issue 4, April 2014, Article No. 44, 2014.

[17] Frederik Gustafsson. Adaptive Filtering and Change Detection. Wiley, October 2000.

[18] Piotr Sobolewski and Michal Wozniak. Concept drift detection and model selection with simulated recurrence and ensembles of statistical detectors. Journal of Universal Computer Science, 19(4):462-483, feb 2013.

[19] Michal Choras, Rafal Kozik, Rafal Renk, Witold Holubowicz. The Concept of Applying Lifelong Learning Paradigm to Cybersecurity. ICIC (3) 2017: 663-671.

Preparation of Response Model to Cyber-attacks at Nuclear Facilities

SAMO TOMAŽIČ & IGOR BERNIK

Abstract Nuclear facilities are a part of a national critical infrastructure and are totally dependent on information technologies [1]. Traditionally, these systems were completely isolated from external networks, but recent transition to digital technology has changed the nature of these systems thus enabling extensive interconnections [2]. The Industrial Control Systems (ICS) are no longer air-gapped as they used to be and consequently much more vulnerable to external threats [3]. Research of various sources (literature, standards, regulations, etc.) has shown that there are several publicly available recommendations for ensuring cybersecurity, but only a few recommendations on response to cyber-attacks at nuclear facilities [4]. Findings are a basis for improvement of what is already available. Understanding of the response to cyber-attacks aimed not only at nuclear, but also at other critical infrastructure facilities; chemical, oil, gas, aviation, etc. The result will be comprehensive model of response to cyber-attacks at nuclear facilities, a model for implementation and a tool for reaching an adequate response at nuclear facilities.

Keywords: • Cyber security • Nuclear security • Critical infrastructure • Cyber-attack • Preparedness • Response •

CORRESPONDENCE ADDRESS: Samo Tomažič, Ministry of environment and spatial planning RS, Slovenian Nuclear Safety Administration, Litostrojska cesta 54, 1000 Ljubljana, Slovenia, e-mail: samo.tomazic@gov.si. Igor Bernik, Ph.D., Associate Professor, University of Maribor, Faculty of Criminal Justice and Security, Kotnikova 8, 1000 Ljubljana, Slovenia, e-mail: igor.bernik@fvv.uni-mb.si.

DOI https://doi.org/10.18690/978-961-286-114-8.4 ISBN 978-961-286-114-8

Available at: http://press.um.si.

1 Introduction

Modern way of life has made each and every one of us totally dependent on information technologies, cyber space and communications. Every day we use our smart devices and we can't imagine our lives without them anymore. These devices are connected to cyber space and to each other [5]. But now days, connection is not limited only to our smart devices, but also to critical infrastructures' digital equipment in energy sector, communication sector, transport sector, etc., which is becoming more and more interconnected [6]. Traditionally, systems like Supervisory Control and Data Acquisition (SCADA), Industrial Control Systems (ICS or I&C), Distributed Control Systems (DCS), etc. were completely isolated from external networks, but recent transition from analog to digital technology, has changed the nature of these systems thus enabling extensive interconnections between ICS, business systems and even extending it to global internet connections. These connections laid the foundation of several different new characteristics of such a digital equipment i.e.; easier maintenance of such an equipment, more functionalities, simple replacement, low cost, etc. [7]. Air-gaps are not fully gone, but now they include one-way guards to mitigate the flow of data between enclaves. There are still nuclear facilities that have air-gapped networks, but we also need to consider that cyber adversaries have developed capabilities for bridging air-gaps using non-cyber domain exploits, what makes ICS now also vulnerable to cyber threats.

This article focuses on digital ICS, used in nuclear sector, mainly at nuclear facilities like nuclear power plants (NPP), and also at research reactors (RR), and storage of radioactive waste. The majority of older nuclear facilities uses analog systems for instrumentation, control and security, but newer ones are focused more on implementation of the digital equipment [8]. During modifications in the older NPPs, digital systems are implemented due to budget cuts, more functionalities, and lack of availability of older analog. But digital systems don't come only with advantages, but also with some weaknesses. One, and also one of the most critical one is, that they are vulnerable and susceptible to cyber-attacks.

In the last decade, the number of cyber-attack has drastically increased [9]. During the research, we have found several cyber-attacks, targeting directly critical infrastructure and also nuclear facilities. The following Fig. 1 represents the timeline on various cyber-attacks on critical infrastructure in the past seven years. Attackers are more and more resourceful at engaging their malicious acts. They are using new ways to attack and they are also focused at nuclear facilities [1].

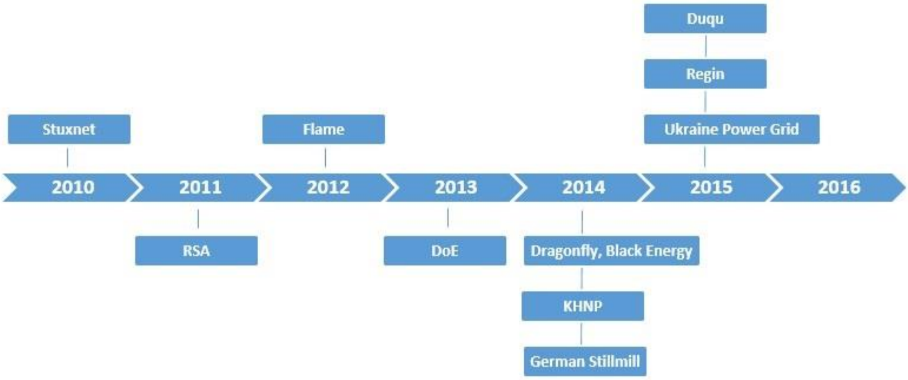


Figure 1: Cyber-attacks directly targeting critical infrastructure.

“Wakeup call” is a term, a lot of scientists, researchers, engineers and other use to describe the Stuxnet [10]. It is a malware, that targeted Siemens ICS, it is very complex, it exploited a long list of vulnerabilities, four of which were 0-days. These vulnerabilities opened an opportunity for malware to self-replicate, update, connect to a remote server, download malicious code and modify the existing code on ICS. Stuxnet had the biggest impact on Iranian uranium enrichment facility in Natanz. The result was about 1.000 destroyed centrifuges, that spun out of control and at the end, thorn themselves apart [11]. Then, there was an attack on the energy sector in the EU and USA in 2014 with Dragonfly, attack on the NATO and EU with BlackEnergy in 2014 [12]. Also in 2014, there was an incident at Korean Hydro and Nuclear Power (KHNP). Hackers used phishing emails to access and download companies’ internal documents. The leakage of documents continued throughout 2015. On December 15th 2015, some of the documents were posted publicly online. The operations of NPPs in Korea were not affected with this cyber-attack [13]. One of the most serious cyber-attacks, that really showed, how vulnerable critical infrastructure actually is, was the attack on Ukraine power grid in 2015. Attackers compromised information systems of three energy distribution companies in the Ukraine and temporarily disrupted electricity supply to approximately 230.000 people for up to 6 hours [14].

Some of these attacks have proven the fact that a remote access is possible even if the systems are isolated or air-gapped [15]. For these attacks to be successful, attackers have to know their target, have sufficient budget, knowledge, motivation and intent [16]. These malicious acts can be carried out by terrorist groups, hacktivists, disgruntled employees, thieves, insiders, or could even be a state sponsored attacks [17]. In cyber domain, we are also familiar with unwilling accomplice, which represents a person, who does a malicious act unknowingly.

We, as a defender, have to be aware that external attackers and insiders are becoming more and more sophisticated, exploits are much easier to get by and vulnerabilities can be found in on-line free databases. Attackers are getting a lot of trainings, they are motivated and also financed by competing organizations, terrorists, states, etc. [18].

Consequences of cyber-attacks can be potentially devastating, just like consequences of natural disasters as earthquake, tsunami, flood, etc. Therefore, we have to ensure that prevention, detection and right response is in place, to mitigate from cyber-attacks, and prevent attackers from accomplishing their goals [19]. Digital systems also have to resilient as much as possible in order to prevent successful cyber-attacks or cascading effects that could result in disruptions or even loss of availability.

In ordinary IT world, there is a lot of written materials on response to cyber-attacks. Organizations like ENISA, NIST and ISO have been publishing all sorts of documents for quite a long time. Some of these documents are guides, some standards, best practices, etc., to help out defenders of ICT systems. A lot of companies, government organizations and others, have adopted these documents and on their basis prepared their own programs, plans, policies, technical guides, etc. So basically, security specialists, and also users in above mentioned organizations, know how to deal with cyber-attacks.

But for us, the most important questions are: how are the nuclear facilities preparing for cyber-attacks, do they know their equipment, equipment's vulnerabilities, possible threats, do they know who is to act during a cyber-attack and how they will act, what procedures are to be used, what to do after the incident, with whom to cooperate during a response, how to work with the law enforcement authorities, how the information is handled by other institutions, etc. When researching various sources and conversations with colleagues from abroad, we found that there is little written and even less implemented at nuclear facilities. Also, it often happens that managers of nuclear facilities and states have a lot of documents prepared, written, but for security reasons, they do not disclose potential good practices with others.

To answer all the above written questions, we have set for ourselves to prepare a comprehensive response model to cyber-attacks at nuclear facilities, developed to such a degree that it can be tested at a real nuclear facility.

Paper presents a theoretical approach and it is a basis for prepared research scenario, which will be implemented at nuclear facilities in Slovenia in the next period.

2 The research

During our extensive research, a lot of various sources, i.e. international standards, guides, best practices, expert articles, doctoral dissertations, literature, regulations, etc., has shown, that there are several publicly available recommendations for ensuring cyber security, but only a few recommendations on response to cyber-attacks at nuclear facilities. Some examples of incident response guides, that focuses on general cyber security are; SANS Incident Handler's Handbook, NIST SP 800-61 Security Computer Security Handling Guide and ENISA Good Practice Guide for Incident Management.

IAEA Computer Security Incident Response Planning at Nuclear Facilities is the only guide that focuses on nuclear facilities. This finding are an opportunity to improve what is already available.

From here on, our research consists of three phases:

- Analysis and preparation of the model (analysis of various sources, interviews with international experts in cyber security and nuclear security, preparation of incident response model).
- Preparation of the exercise scenario and criteria (organization of the national exercise, preparing the scenario, injects and the criteria by which we verify the efficiency of the model in practice).
- Validation and verification of the model during the national exercise.

This research is an important contribution to a better understanding of cyber security in other infrastructure facilities like chemical, oil, gas, etc., that were built for a relatively long period of time. Managing these types of facilities requires a lot of specific knowledge and expertise. The result of the research is a comprehensive response model to cyber-attacks at nuclear facilities, a model, which when implemented, is also a tool for achieving a good protection of a particular nuclear facility.

2.1 Analysis and preparation of the model

The research is based on a collection of qualitative data. A basis to identify what has actually been done and where there is still a room for improvement is descriptive analysis of various sources.

Proposed interview questions are based on descriptive analysis of all the above listed sources. A sufficient number of international experts in nuclear security and cyber security need to be selected as interviewees. Proposed focus group consists of international experts, employees at nuclear facilities, employees at regulatory bodies and other international experts in charge of cyber security at nuclear facilities.

The results of the interviews are an insight into an actual situation at nuclear facilities around the world, as well as possible improvements to the current shortcomings.

Below are listed three examples of possible interview questions:

- How and who in your country participates in law, legislation and regulation making regarding cyber security at nuclear facilities?

The answer to this question tells us whether the member state actually has a law, legislation or regulation related to cyber security. If regulators, operators and other relevant stakeholders are cooperating and working together to improve cyber security.

- Do you have a threat assessment or Design Basis Threat (DBT) related to cyber security? Do you have a program or plan for cyber security?

Implementing cyber security at nuclear and other infrastructure facilities could represent a really difficult task for the operators and also for the state organizations. The answer to this question tells us where the member state is at the moment and what are the biggest issues it deals with.

- Do you have examples of best practices in cyber security in your country, that you would like to highlight?

With this answer, we might get some ideas on how to better prepare the model, and what are lessons learned by other, so we don't make the same mistakes.

Based on a grounded theory [20], more and more data has to be collected. Later on, these data should be grouped into concepts, and then into categories, which may become the basis for a new theory. This is the starting point for preparation of the response model to cyber-attacks at nuclear facilities.

2.2 Preparation of the exercise scenario and criteria's

National exercise shows us how the model works in real life. This type of exercise is the best way to find pros and cons of the prepared model. To make the exercise work, the permission of the managements of all stakeholders has to be obtained. During security exercises in nuclear sector in Slovenia, stakeholders include, but are not limited to, staff from the NPP, Ministry of Interior Affairs, Ministry of Defense and Slovenian Nuclear Safety Administration. But when cyber-elements are included in an exercise, also other cyber security relevant organization's like national CERT, Ministry of Public Administration and others need to participate.

According to a long-standing practice, the organization of the exercise won't represent the biggest obstacle to overcome. The most critical obstacle is the preparation of a meaningful possible scenario and appropriate injects for the exercise. According to [21], preparing a good scenario, is a great challenge. A plausible cyber-attack has to be selected, and the effects and impact of a selected cyber-attack to the NPP, its systems and components determined, and only then a response to a specific type of cyber-attack can be planned.

Using prewritten criteria's, the efficiency of the model in practice is verified. Criteria's is prepared simultaneously with the preparation of scenario. In this way, it can be verified, whether the developed model is successful or where improvements of the model are needed.

2.3 Validation and verification of the model during the national exercise

Safety and security exercises are really common in nuclear sector. Every country that has a nuclear power plant or even just a small nuclear facility, has an incident management response plan in place. Unfortunately, those plans most of the time don't include a cyber-element. So this is a great opportunity for testing the quality of a model in practice. With validation and verification, we can check, if the developed model meets all the requirements and that it fulfills its intended purpose. During validation phase, participating stakeholders can accept the model as suitable, or give a recommendation on how to improve it. The verification is therefore a feedback on whether the model meets and complies with corresponding regulations and requirements.

3 The model

Developed response model to cyber-attacks at nuclear facilities consists of four steps; (1) preparation for a cyber-attack, (2) detection of a cyber-attack or an intruder in the system, (3) response to a cyber-attack and (4) analysis of the incident or event.

During the research, additional knowledge and already existing best practices, might change the view on main steps taken, and parts, tasks and phases could be rearranged, added or deleted during the development of the model.

3.1 Preparation

Preparation is the most essential step. Of course, we cannot prepare for everything and we also cannot think of every malicious act there is out there, but with good preparation we can minimize the window of opportunity for the cyber adversaries to perform their malicious acts.

We can do that by defining the baseline of each critical system (normal operation), regularly following the new and emerging threats, assembling the right Computer Security Incident Response Team (CSIRT), allocating the right equipment and task to the CSIRT team, predefine alarming and calling of personnel, reporting the right information to the right people, organizing facilities or incident response centers, regularly attend different types of trainings (awareness or specialized) and organize announced and unannounced exercises for all relevant stakeholders.

3.2 Detection

The most important part of detection is to ensure that monitoring is in place. Monitoring can be done automatically or manually. Thus, if we want to detect any anomaly on time, we have to have a dedicated equipment and personnel with the right knowledge in place. After the detection, we have to distinguish between normal and abnormal operation and respond in the predefined manner. Just like in safety emergency preparedness, escalation levels should be in place.

3.3 Response

When cyber-attack has been detected, all focus has to be directed in to the right response with the goal to put the nuclear facility back into normal operation as soon as possible. Cyber-attack has to be contained, infection eradicated and all the systems have to be recovered and put back into normal operation. If evidence collection, destroyed equipment, etc. requires, we also have to reinstall software and physically change the equipment. But during this step, it is essential, we don't forget about evidence collection. This is usually done by the law enforcement agencies. Evidence can be later on used to catch the cyber adversaries, so similar cyber-attack won't repeat after putting the facility back into normal operation.

3.4 Analysis

Analysis is crucial and brainstorming after the cyber-attack is a must. There is a lot we can learn from an incident, events and even from our own mistakes [22]. There are several databases available for operators and regulators, where different stakeholders in industry and government change information. Also inputs like suggestions and recommendations from involved people are desirable. After all, we are in a continuous process, where we continue to improve ourselves, equipment and procedures we use.

4 Challenges

During the research of nuclear sector and cyber security within, we have identified some obstacles, that we have to overcome while preparing the response model to cyber-attacks at nuclear facilities. Some of those obstacles already occurred during the research:

- Information sharing is the biggest issue. Stakeholders in nuclear sector are really cautious while sharing any information. The main reason is, if confidential information falls into the wrong hands, it could represent a serious problem not only for security, but also for safety.
- Sharing operational experiences in safety domain is more or less common, but cyber incidents or any kind of experiences in cyber at nuclear facilities, are not always publicly shared, so we can't learn from others.

In the phase of preparation of the model, the following additional obstacles might occur:

- Decision on how to build a response model. There are two possibilities. One is to create self-standing response model, and the second one is to implement the developed model into an existing emergency preparedness plan.
- Assembling the right CSIRT. Usually not everyone is the right person for the job. The model has to take into account the right decision making to include the most suitable person.
- Creating real-life attack scenario and injects, which helps all experts to stay sharp, coordinated and ready to do the right thing.

- Getting all the stakeholders to work together and to share information, could represent a major obstacle.

Deep dive into publicly available actual events, vulnerabilities, threats, etc., is a way to overcome obstacles during the research. A lot of team work and engagement from participants and the management also contributes to achieving the set goal.

5 Discussion

Protection of a national critical infrastructure is one of the most important tasks for the state. With protecting, we are now days focused not only the physical protection, but also on protecting the cyber space. There are many ways of accomplishing this, some of those are preparing strategies, laws, regulations, cooperation on various levels with relevant stakeholders, etc.

In regards to nuclear sector, which is also a part of a national critical infrastructure, we have done quite a lot. In Slovenia, the key organization for ensuring nuclear safety and also cyber security at nuclear facilities, is Slovenian Nuclear Safety Administration (SNSA), which is one of the nuclear regulators in the country. In the past several years, SNSA has worked on laws, prepared regulations, organized national and international trainings and also established a national working group on cyber security at nuclear facilities. This group consists of all relevant stakeholders in cyber security in the country and its main purpose is sharing information and cooperation.

During a continued research a full support, not only from the SNSA, but also from other stakeholders from the above mentioned national working group is needed. By support, we mean cooperation at interviews, giving advices on plausible scenarios and criteria's, preparing and executing the exercise and the final evaluation of response model to cyber-attacks at nuclear facilities. Preparing the model requires a lot of cooperation and patience, because some actions tend to realize really slowly, especially if additional approvals and double checking is required.

Some member states have already prepared and published their cyber security regulations for nuclear facilities, some are in process of making them and some are thinking about it. But all of them have one thing in common. The majority of published regulations already include provisions for incident response, and the future ones will have to. In Slovenia, SNSA has already published cyber security regulations for nuclear facilities, which are adopted mainly from US NRC 10 CFR 73.54 Protection of digital computer and communication systems and networks [23]. These and other similar regulations state, that operators have to include measures for incident response and recovery for cyber-attacks. But without detailed and structured guides, operators all over the world are struggling. Therefore, a response model to cyber-attacks at nuclear facilities is helpful for the operators, not only in Slovenia, but also in other member states. Especially for the ones that are at the beginning of their cyber program development. The model is also a great addition for the ones that already have plans, procedures, guides, etc. on incident response to cyber-attacks.

6 Conclusions and further research

Review of domestic and international literature has led us to conclude, there is not a lot of publicly available sources, which operators of nuclear facilities and regulators could use, to better prepare themselves to respond to cyber-attacks. But, there are a lot of sources, that have proven, that the threat is real, it is out there, and it is just waiting for us to make a mistake. There are also several examples of cyber-attacks pointed against critical infrastructure, energy sector and within, the nuclear sector. The "wakeup call", as the researchers and experts call the Stuxnet, changed our opinion on "totally secure" air-gaped systems.

Research also showed us, there is an urgent need for planning a structured and well organized incident response plan, that helps all stakeholders in nuclear sector to better prepare, detect and respond to cyber-attacks at nuclear facilities.

References

[1] C. Baylon, R. Brunt and D. Livingstone. 2015. Cyber Security at Civil Nuclear Facilities: Understanding the Risks. London, UK: Chatham House Report.

[2] P. Litherland, R. Orr and R. Piggin. 2016. Cyber security of operational technology: Understanding differences and achieving balance between nuclear safety and nuclear security. 2016 World Conference on Innovation, Engineering, and Technology (IET 2016). 24-26.6.2016. Sapporo: IET

[3] K.K. Green. 2014. Modeling and Testing a Cyber Secure Protection System for Smart Grid. Washington, DC: Howard University.

[4] D.D. Dudenhoeffer, K. Mrabit, J. Hilliard and M.T. Rowland. 2015. Gates, guards, guns and geeks: the changing face of nuclear security and the IAEA's leading role in promoting computer security for nuclear facilities. Nuclear Plant Instrumentation, Control and Human-Machine Interface Technologies (NPIC & HMIT 2015), 22-26.2.2015. Charlotte: American Nuclear Society.

[5] T. Macaulay. 2017. RIoT Control: Understanding and Managing Risks and the Internet of Things. Cambridge: Morgan Kaufmann.

[6] M.T. Rowland, D.D. Dudenhoeffer and J.S. Purvis. 2017. Computer Security for I&C Systems at Nuclear Facilities. Nuclear Plant Instrumentation, Control and Human-Machine Interface Technologies (NPIC & HMIT 2017), 11-15.6.2017. San Francisco: American Nuclear Society.

[7] M. Ficco, M. Choraś and R. Kozik. In print. Simulation platform for cyber-security and vulnerability analysis of critical infrastructures. Journal of Computational Science.

[8] J. Park, Y. Suh and C. Park. 2016. Implementation of cyber security for safety systems of nuclear facilities. Progress in Nuclear Energy, 88, 88-94.

[9] I. Onyeji, M. Bazilian and C. Bronk. 2014. Cyber Security and Critical Energy Infrastructure. The Electricity Journal volume 27(2), 52-60.

[10] J.F. Brenner. 2013. Eyes wide shut: The growing threat of cyber attacks on industrial control systems. Bulletin of the Atomic Scientists, 69(5), 15-20.

[11] N. Falliere, L. O. Murchu and E. Chien. 2011. W32.Stuxnet.Dossier, Version 1.4 (February 2011). Cupertino, CA: Symantec Corporation.

[12] Symantec Security Response. 2014. Dragonfly: Cyberespionage Attacks Against Energy Suppliers, 1.21 (July 7). Cupertino, CA: Symantec Corporation.

[13] K.-B. Lee and J.-I. Lim. 2016. The Reality and Response of Cyber Threats to Critical Infrastructure: A Case Study of the Cyber-terror Attack on the Korea Hydro & Nuclear Power Co., Ltd. KSII Transactions on Internet and Information Systems, Vol 10, 2 (Feb. 2016), 857 – 880.

[14] M. R. Lee, J. M. Assante and T. Conway. 2016. Analysis of the Cyber Attack on the Ukrainian Power Grid. Washington, DC: E-ISAC.

[15] J.R. Lindsay. 2013. Stuxnet and the Limits of Cyber Warfare. London:Routledge

[16] R. Masood. 2016. Assessment of Cyber Security Challenges in Nuclear Power Plants. Washington, DC: Cyber Security and Privacy Research Institute of the George Washington University.

[17] J. Xie, A. Stefanov and C. Liu. 2016. Physical and cyber security in a smart grid environment. Wiley Interdisciplinary Reviews: Energy and Environment, 5(5), 519-542.

[18] M. Nalabandian, A, Van Dine and P. Stoutland. 2016. Global action on cybersecurity at nuclear facilities: Moving beyond the status quo. Washington, DC: NTI - Nuclear Threat Initiative.

[19] Computer security incident response planning at nuclear facilities. 2016. Vienna: International Atomic Energy Agency.

[20] D. Remenyi. 2014. Grounded Theory, 2 (Jul. 2014). Reading, UK: Academic Conferences and Publishing International Limited.

[21] J.M. Hollern and P.F. Stringfellow. 2015. Considerations for integrating cyber security requirements into the nuclear facility emergency preparedness plan. NPIC and HMIT 2015, Vol 3, 1928 – 1935.

[22] M.J. West-Brown, D. Stikvoort, K.P. Kossakowski, G. Killcrece, R. Ruefle and M. Zajicek. 2003. Handbook for Computer Security Incident Response Teams (CSIRTs). Pittsburgh: Carnegie Mellon University.

[23] Rules on radiation and nuclear safety factors. 2016. Ljubljana: Slovenian Nuclear Safety Administration.

Use of Mobile Devices in Hospitals and Perceived Data Breach Consequences: An Explorative Study

SIMON VRHOVEC & BLAŽ MARKELJ

Abstract Health care professionals are increasingly using mobile devices in their everyday work to improve patient care. Hospitals may however fail to adequately address the use of mobile devices and adapt their information security policies in time. Health care professionals may use both their personal and work mobile devices for their everyday work. Sometimes they do it without adhering to an adequate hospital information security policy. The objective of this paper is to study the relation between the use of mobile devices, adhering to hospital information security policy and perceived consequences of data breaches. An exploratory survey (N = 95) has been conducted in a Slovenian hospital. Respondents were asked about the use of their personal and work mobile devices for accessing medical data, adhering to the hospital information security policy, and the perceived consequences of data breaches for themselves, the hospital and the patients. The results show that perceived personal consequences are negatively correlated with personal and work mobile device use for work. Also, adhering to information security policy is positively correlated with perceived data breach consequences for both the patients and the hospital.

Keywords: • Hospital • mobile devices • information security • patient privacy • health care •

CORRESPONDENCE ADDRESS: Simon Vrhovec, Ph.D., Assistant Professor, University of Maribor, Faculty of Criminal Justice and Security, Kotnikova 8, 1000 Ljubljana, Slovenia, e-mail: simon.vrhovec@fvv.uni-mb.si. Blaž Markelj, Ph.D., Assistant Professor, University of Maribor, Faculty of Criminal Justice and Security, Kotnikova 8, 1000 Ljubljana, Slovenia, e-mail: blaz.markelj@fvv.uni-mb.si.

DOI https://doi.org/10.18690/978-961-286-114-8.5 ISBN 978-961-286-114-8

Available at: http://press.um.si.

1 Introduction

Hospitals are introducing mobile devices into everyday work of health care professionals to improve patient care and their work processes [1, 14, 15]. Mobile devices are just as ubiquitous in healthcare as they are in general and health care professionals can use both their work or personal mobile devices for everyday work [15, 19, 20]. Using personal mobile devices at work (*bring-your-own-device*, BYOD) is often preferred by hospitals as it enables them to significantly lower the costs needed to equip all personnel with mobile devices that are used only in hospital settings [1, 7, 8, 14].

Mobile device adoption by health care professionals however brings new information security issues as incidents related to mobile devices account for most data breaches in health care in recent years [4]. Hospitals are therefore required to adapt the hospital information policies to encompass both work and personal mobile device use and enforce them [1, 8, 14]. Ensuring that health care professionals adhere to the hospital information security policy may however prove to be quite challenging to achieve [9, 16, 19].

The objective of this paper is to study the relation between the use of mobile devices for accessing medical data, adhering to hospital information security policy and perceived consequences of potential data breaches. To achieve this, we conducted an explorative survey in a Slovenian hospital. Respondents were asked about the use of their personal and work mobile devices for accessing medical data, adhering to the hospital information security policy, and the perceived consequences of data breaches for themselves, the hospital and the patients.

2 Background

Mobile device use in hospital settings offers a variety of new possibilities [15]. Health care professionals may use them as an alternative to workstations for accessing medical data which enables them to access it from wherever needed, the patient room, a meeting, the patient's home or elsewhere [10, 17]. Use of mobile devices tends to increase work satisfaction of health care professionals and improves direct communication between them and the patients [10, 18].

Mobile devices may be costly for hospitals to introduce and maintain [1]. To lower the costs, hospitals may encourage the use of personal mobile devices at work [1, 12]. BYOD is also convenient for health care professionals as they are already familiar with the device making it a win-win situation [19]. Not everything is good about BYOD though. Accessing medical data from a device that is used for both personal and work use is a security issue *per se*. This is a fundamental trade-off between data security and data access [2]. The hospital has few means to control the cybersecurity of the mobile device, e.g., by checking for VPN connection or disabling access for rooted or jailbroken devices [1].

The hospital information security policy needs to be adapted to the de facto use of mobile devices in the hospital. Research shows that over 90 percent of health care professionals use their own mobile devices at work to access medical data [4, 12]. However, only 38 percent of hospitals define a formal policy of mobile device use [12, 17, 18]. Low awareness of cybersecurity threats and hospital information security policies of health care professionals seem to be a major challenge [19]. Despite attempts to ensure information security and patient privacy, most of recent data breaches seem to be related to mobile devices [4].

We can distinguish consequences of data breaches on three levels. First, a medical data breach directly affects the patients whose data has been exposed. Medical data can be used to acquire direct financial gain, commit an electronic fraud, steal the medical identity, or extort the victims [17]. Medical identity theft is wide-spread and can have severe financial and medical consequences if patient's medical record is contaminated with medical data of a third person [4, 13]. Therefore, we develop the first set of hypotheses:

> **H1a**. Perceived consequences of a data breach for the patients will be correlated with work mobile device use.
>
> **H1b**. Perceived consequences of a data breach for the patients will be correlated with personal mobile device use.
>
> **H1c**. Perceived consequences of a data breach for the patients will be correlated with adhering to the hospital information security policy.

Second, the data breach may affect the hospital where the breach occurred. Patients trust hospitals with their most private and private information and hospitals seek to keep their reputation as trustworthy organizations [4]. Recovering the lost reputation is a tough job [4]. In addition to losing patients, hospitals face high fines and lawsuits from patients [4]. These arguments suggest the second set of hypotheses:

> **H2a**. Perceived consequences of a data breach for the hospital will be correlated with work mobile device use.
>
> **H2b**. Perceived consequences of a data breach for the hospital will be correlated with personal mobile device use.
>
> **H2c**. Perceived consequences of a data breach for the hospital will be correlated with adhering to the hospital information security policy.

Third, the data breach may affect the person responsible for it, i.e., the health care professional using the mobile device at the time of the data breach. A data breach may significantly affect a person's career or there may be no consequences at all depending on the hospital policy. These arguments support the following hypotheses:

> **H3a**. Perceived personal consequences of a data breach will be correlated with work mobile device use.

H3b. Perceived personal consequences of a data breach will be correlated with personal mobile device use.

H3c. Perceived personal consequences of a data breach will be correlated with adhering to the hospital information security policy.

3 Methods

To test our model, we conducted an exploratory survey in a Slovenian hospital. The survey was conducted among health care professionals participating in information security training organized by a third-party cybersecurity company. Randomly chosen groups of attendants were asked to complete the survey before attending the training. In total, 150 surveys have been administered and 95 respondents returned the survey representing a response rate of 63 percent. The number of missing values ranged from 1.1 to 3.2 percent except for work device use which had 7 missing cases (7.4 percent).

All constructs were measured by using single-items as a very high degree of parsimony was required [5, 11]. Under specific conditions, the predictive validity of single-item measures is comparable to the predictive validity of multi-item measures [3, 6, 11]. Due to the exploratory nature of the study, the use of single-item measures thus seems reasonable. The survey items are presented in the Appendix A.

Table 1: Demographic characteristics

Characteristic	N	Percent
Health care professional		
Physician	2	2.11
Nurse	77	81.05
Administration personnel	9	9.47
Not specified	7	7.37
Gender		
Male	16	16.8
Female	77	81.1
Not specified	2	2.1

Demographics of the respondents are presented in Table 1.

4 Results

Table 2 includes means and standard deviations (SD) for all studied constructs.

Table 2: Descriptive statistics

Construct	Mean	SD
1. Perceived data breach consequences for patients	6.56	0.811
2. Perceived data breach consequences for hospital	6.45	0.863
3. Perceived personal consequences of a data breach	6.65	0.617
4. Work mobile device use	2.77	1.849
5. Personal mobile device use	1.60	1.177
6. Adhering to information security policy	5.62	1.146

Since some constructs did not follow a normal distribution, we calculated Kendall's Tau non-parametric rank correlation coefficients between them. Table 3 includes inter-correlations for all studied constructs. The results support hypotheses H2c and H3b (p < 0.01) and hypotheses H1c and H3a (p < 0.05). Other results are nonsignificant and do not support the remaining hypotheses H1a, H1b, H2a, H2b and H3c.

Table 3: Correlation matrix

Construct	1	2	3	4	5
2	0.475^{***}				
3	0.364^{***}	0.257^{**}			
4	-0.166	-0.106	-0.197^{*}		
5	-0.103	-0.142	-0.268^{**}	0.321^{**}	
6	0.219^{*}	0.249^{**}	0.041	-0.128	-0.080

* p < 0.05;** p < 0.01; *** p < 0.001.

The results of hypotheses testing are presented in Fig. 1. In addition to the test of our hypotheses, it includes other results of interest.

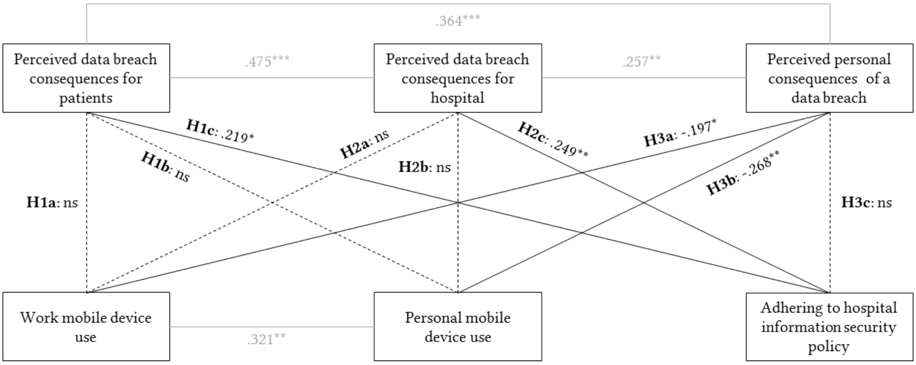

.364***
Perceived data breach consequences for patients
.475***
Perceived data breach consequences for hospital
.257**
Perceived personal consequences of a data breach
H1c: .219*
H2a: ns
H2c: .249**
H3a: -.197*
H1b: ns
H2b: ns
H3b: -.268**
H1a: ns
H3c: ns
Work mobile device use
.321**
Personal mobile device use
Adhering to hospital information security policy

Figure 1: Hypotheses testing results.

First, there are significant correlations between perceived personal consequences of a data breach and perceived data breach consequences for hospital and patients (p < 0.001). Next, the correlation is also significant between perceived data breach consequences for hospital and patients (p < 0.01). Finally, there is a significant correlation between work and personal mobile device use (p < 0.01).

5 Discussion

The results of this explorative study support only 4 out of the 9 developed hypotheses. The confirmation of hypotheses H1c and H2c shows the importance of perceived data breach consequences for the hospital and the patients in adhering to the hospital information security policy. The higher the perceived data breach consequences for the hospital and the patients the higher the adherence to the hospital information security policy. It is however surprising that hypothesis H3c has not been confirmed as it suggests that the perceived personal consequences of a data breach do not play an important role in adhering to the hospital information security policy. This suggests that health care professionals do not take the hospital information security policy for their own despite the relatively high scores of this construct (5.62 ± 1.146). In other words, these findings suggest that the health care professionals do not identify completely with the hospital regarding its information security policy. Nevertheless, the health care professionals respect the need to protect the hospital and the patients from data breaches by adhering to the hospital information security policy.

The confirmation of hypotheses H3a and H3b and non-confirmation of the remaining hypotheses H1a, H1b, H2a and H2b suggests quite the opposite for mobile device use. Higher perceived personal consequences of a data breach seem to hinder the adoption of mobile devices. This may be attributed to the lack of a clear hospital information security policy on mobile devices as the hospital had no policy for personal mobile devices and work mobile devices were only being formally introduced. Perceived data breach consequences for the hospital or the patients however do not seem to play an important role in the use of mobile devices by health care professionals. This is quite surprising as it suggests that health care professionals do not consider the consequences that the use of their mobile device may have for either the hospital or the patients. The difference seems to vary depending on the type of mobile device. Work mobile device use is more loosely correlated with perceived personal consequences of a data breach than personal mobile device use. Also, nonsignificant correlations related to work mobile device use seem to be closer to the significant one than the nonsignificant correlations related to personal mobile device use. This gives us a hint that health care professionals could only loosely relate the use of their personal mobile devices to data breach consequences for hospitals and patients. Again, this could still be attributed to the lack of a clear hospital information security policy on mobile devices.

There are strong correlations between all three perceived data breach consequences. These results suggest that health care professionals relate their personal consequences more to the consequences for the patients than those for the hospital. This shows that the health care professionals first think of their patients and only after of their employer. In a way, this supports the finding that they do not identify well with the hospital regarding its information security policy.

The use of work and personal mobile devices is relatively low (2.77 ± 1.849 and 1.60 ± 1.177, respectively) which can be attributed to the fact that the mobile devices are just being introduced. There is a strong correlation between work and personal mobile device use. This suggests that health care professionals could use personal mobile devices for work only after they have started using the work mobile devices. Another explanation would be that they are even able to use personal mobile devices for work after the hospital information system adds the mobile device access functionality and they become aware of it. This highlights the importance of covering both work and personal mobile device use in the hospital information security policy.

As with all research, the reader should consider the limitations of this study when interpreting its results. First, the survey was done in a single subject organization. The studied hospital could be considered as a typical hospital in the process of introducing mobile devices to its processes. Nevertheless, the reader should be cautious when generalizing the findings of this study. Further research should aim to include more hospitals from different cultural contexts and different stages of mobile device adoption. Second, the use of single-item measures in the survey has its drawbacks as it does not allow reliability and validity analysis of the survey instrument. Further research should aim to use multi-item measures which allow rigorous testing of the survey instrument. Third, the respondents may not have had the same notion of the information security policy and a data breach entail as they were not provided any explanation prior to the survey. It may thus be possible that the responses were not indicative of the same concept. Further research should consider focusing on clear and real scenarios when preparing new items. Fourth, the subjects of the study were all health care professionals. A study comparing different health care professionals (e.g., administration personnel, physicians and nurses) would provide useful insights into the differences between them. Fifth, it would be also useful to incorporate level of experience in the healthcare domain as a control variable in the analysis.

6 Conclusion

This paper reports on an explorative study on the use of mobile devices in hospitals and perceived data breach consequences. The health care professionals consider data breach consequences for the hospital and the patients when adhering to the hospital information security policy. They however do not relate data breach consequences for themselves to adhering to the hospital information security policy suggesting they do not see personal benefits in it. Hospitals need to improve this by either better promoting its information security policy among health care professionals or updating it in a way that they would see benefits in it.

When using mobile devices, health care professionals consider only data breach consequences for themselves. They do not consider data breach consequences for the hospital or the patients which suggests that they do not relate their mobile device, either work or personal, to it. Hospitals may improve this lack of awareness by raising it. The health care professionals need to understand that the use of mobile devices can affect both the hospital and the patients in the case a data breach happens. At the same time, they need to adequately understand the risks of using mobile devices and measures to tackle them in order not to hinder mobile devices adoption in hospitals.

A survey items

All items were scored on a seven-point Likert scale from 1 (*I strongly disagree*) to 7 (*I strongly agree*).

Perceived data breach consequences for patients
Data breaches are very harmful for the affected patients.

Perceived data breach consequences for hospital
Data breaches are very harmful for the hospital.

Perceived personal consequences of a data breach
Data breaches are very harmful for the one responsible.

Work mobile device use
I use my work mobile device to access patient data very often.

Personal mobile device use
I use my personal mobile device to access patient data very often.

Adhering to information security policy
I always adhere to the hospital information security policy.

References

[1] Al Ayubi, S.U., Pelletier, A., Sunthara, G., Gujral, N., Mittal, V. and Bourgeois, F.C. 2016. A Mobile App Development Guideline for Hospital Settings: Maximizing the Use of and Minimizing the Security Risks of "Bring Your Own Devices" Policies. JMIR mHealth and uHealth. 4, 2 (2016), e50. DOI:https://doi.org/10.2196/mhealth.4424.

[2] Bai, G., Jiang, J. (Xuefeng) and Flasher, R. 2017. Hospital Risk of Data Breaches. JAMA Internal Medicine. 177, 6 (Jun. 2017), 878. DOI:https://doi.org/10.1001/jamainternmed.2017.0336.

[3] Bergkvist, L. and Rossiter, J.R. 2007. The Predictive Validity of Multiple-Item Versus Single-Item Measures of the Same Constructs. Journal of Marketing Research. 44, 2 (May 2007), 175–184. DOI:https://doi.org/10.1509/jmkr.44.2.175.

[4] Bitglass 2014. The 2014 Bitglass Healthcare Breach Report.

[5] Bunderson, J.S. and Boumgarden, P. 2010. Structure and Learning in Self-Managed Teams: Why "Bureaucratic" Teams Can Be Better Learners. Organization Science. 21, 3 (2010), 609–624. DOI:https://doi.org/10.1287/orsc.1090.0483.

[6] Diamantopoulos, A., Sarstedt, M., Fuchs, C., Wilczynski, P. and Kaiser, S. 2012. Guidelines for choosing between multi-item and single-item scales for construct measurement: a predictive validity perspective. Journal of the Academy of Marketing Science. 40, 3 (May 2012), 434–449. DOI:https://doi.org/10.1007/s11747-011-0300-3.

[7] Ehrler, F., Blondon, K., Baillon-Bigotte, D. and Lovis, C. 2017. Smartphones to Access to Patient Data in Hospital Settings: Authentication Solutions for Shared Devices. 14th

International Conference on Wearable Micro and Nano Technologies for Personalized Health, pHealth 2017 (Eindhoven, The Netherlands, 2017), 73–78.

[8] Faulds, M.C., Bauchmuller, K., Miller, D., Rosser, J.H., Shuker, K., Wrench, I., Wilson, P. and Mills, G.H. 2016. The feasibility of using "bring your own device" (BYOD) technology for electronic data capture in multicentre medical audit and research. Anaesthesia. 71, 1 (Jan. 2016), 58–66. DOI:https://doi.org/10.1111/anae.13268.

[9] Giles-Smith, L., Spencer, A., Shaw, C., Porter, C. and Lobchuk, M. 2017. A Study of the Impact of an Educational Intervention on Nurse Attitudes and Behaviours toward Mobile Device Use in Hospital Settings. Journal of the Canadian Health Libraries Association / Journal de l'Association des bibliothèques de la santé du Canada. 38, 1 (2017), 0–2. DOI:https://doi.org/10.5596/c17-003.

[10] HIMSS Analytics 2014. 2014 Mobile Devices Study.

[11] Lee, H., Delene, L.M., Bunda, M.A. and Kim, C. 2000. Methods of Measuring Health-Care Service Quality. Journal of Business Research. 48, 3 (Jun. 2000), 233–246. DOI:https://doi.org/10.1016/S0148-2963(98)00089-7.

[12] Martínez-Pérez, B., de la Torre-Díez, I. and López-Coronado, M. 2015. Privacy and Security in Mobile Health Apps: A Review and Recommendations. Journal of Medical Systems. 39, 1 (2015), 181: 1-8. DOI:https://doi.org/10.1007/s10916-014-0181-3.

[13] McDavid, J.P. 2013. HIPAA Risk Is Contagious: Practical Tips to Prevent Breach. The Journal of Medical Practice Management. 29, 1 (2013), 53–55.

[14] Motulsky, A., Wong, J., Cordeau, J.-P., Pomalaza, J., Barkun, J. and Tamblyn, R. 2016. Using mobile devices for inpatient rounding and handoffs: an innovative application developed and rapidly adopted by clinicians in a pediatric hospital. Journal of the American Medical Informatics Association. 10, 2 (Aug. 2016), ocw107. DOI:https://doi.org/10.1093/jamia/ocw107.

[15] Sharpe, B. and Hemsley, B. 2016. Improving nurse-patient communication with patients with communication impairments: Hospital nurses' views on the feasibility of using mobile communication technologies. Applied Nursing Research. 30, (2016), 228–236. DOI:https://doi.org/10.1016/j.apnr.2015.11.012.

[16] Sher, M.-L., Talley, P.C., Cheng, T.-J. and Kuo, K.-M. 2017. How can hospitals better protect the privacy of electronic medical records? Perspectives from staff members of health information management departments. Health Information Management Journal. 46, 2 (May 2017), 87–95. DOI:https://doi.org/10.1177/1833358316671264.

[17] Storbrauck, L. 2015. Mobile Device Use: Increasing Privacy and Security Awareness for Nurse Practitioners.

[18] The Office of the National Coordinator for Health Information Technology 2015. Guide to Privacy and Security of Electronic Health Information.

[19] Vrhovec, S.L.R. 2016. Challenges of mobile device use in healthcare. 39th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO 2016) (Opatija, Croatia, 2016).

[20] Whipple, E.C., Allgood, K.L. and Larue, E.M. 2012. Third-year medical students' knowledge of privacy and security issues concerning mobile devices. Medical Teacher. 34, 8 (2012), e532–e548. DOI:https://doi.org/10.3109/0142159X.2012.670319.

University of Maribor Press

Mobile Security: Two Generations of Potential Victims

BLAŽ MARKELJ & SABINA ZGAGA

Abstract Paper presents the results of two research studies, conducted among two generations, i.e. students and employees in organisations, with the aim of generating a comprehensive overview of their use of mobile devices and observing the inter-generational differences. The research studies also served as a basis for determining whether users of mobile devices are at risk of becoming victims of threats to information security. The purpose of the second (theoretical) part of this paper is therefore to present the status of victims of such security incidents.

The boundary between personal and private use of mobile devices has disappeared. The lack of knowledge and rules concerning the use of mobile devices represents the main problem. Victims of threats to mobile devices information security in Slovenia have certain possibilities at their disposal to protect their interests. The problem stems from the gap between the need to react quickly in order to protect one's personal rights and the lengthy legal proceedings.

The value of this paper is in combination of information security and legal aspects and in comparison of the results of two research studies, conducted among two generations.

Keywords: • information security • mobile devices • criminal act • encroachment • personal dignity •

CORRESPONDENCE ADDRESS: Blaž Markelj, Ph.D., Assistant Professor, University of Maribor, Faculty of Criminal Justice and Security, Kotnikova 8, 1000 Ljubljana, Slovenia, e-mail: blaz.markelj@fvv.uni-mb.si. Sabina Zgaga, Ph.D., Constitutional Court, Beethovnova ulica 10, 1001 Ljubljana, Slovenia, e-mail: sabinazgaga@gmail.com.

DOI https://doi.org/10.18690/978-961-286-114-8.6 ISBN 978-961-286-114-8

Available at: http://press.um.si.

1 Introduction

Our society depends on the use of information and communication technology in every aspect of its activity. Mobile devices have become an important part of such technology and have lately literally over flooded the users. According to IDC [1], there was a 2.5 percent growth in year 2016 in mobile device sales. Moreover they are expecting 3 percent growth in the year 2017, because of various new devices and that in the year 2018, this percentage would jump to 4.5 percent and that would mean more potential victims to cybercrime. Mobile devices represent an important factor of daily decision-making, as they enable quick access to information and different ways of communication and are important for keeping the relations between companies and their clients strong [2]. However, users are not sufficiently aware of the disadvantages of mobile devices and threats to mobile security, which should certainly not be disregarded. Daily news and reports demonstrate that threats to mobile devices are the most rapidly growing threats to information security in general. Mcafee [3, 4, 5, 6] noticed, that the trends of threats to mobile devices are continuously growing in the years 2014, 2015, 2016 and 2017. The authors of this paper conducted two research studies among two generations, i.e. students and employees in organisations, with the aim of generating a comprehensive overview of their use of mobile devices. Accordingly, they observed inter-generational differences regarding the use of mobile devices. Furthermore, today's students will soon become employees of various organisations. It is therefore necessary to analyse the use of mobile devices among students in a comprehensive manner. The current state-of-play in organisations is also important, since they are operating with sensitive data on daily basis.

The aforementioned research studies also served as a basis for determining whether users of mobile devices are at risk of becoming victims of threats to information security. Such security incidents could represent a major encroachment upon personal rights of victims-users of mobile devices, particularly upon their right to privacy. This is why the second part of this paper focuses on victims of such security incidents and their status according to the Slovene law. The paper primarily refers to the victims' interest in the prevention and cessation of any further encroachment upon their personal rights as provided by the Code of Obligations [7]. However, victims may also request an interim order for a temporary injunction to prevent further encroachment upon their rights. Since the materialisation of threats to the information security of mobile devices may simultaneously represent the realisation of the legal elements of a specific criminal offence, e.g. abuse of personal data according to the Slovene Criminal Code (2017) [8], this paper also deals with the status of victims of such security incidents in the framework of criminal law. This is applicable particularly when victims are considered injured parties, as that they have the possibility and duty to file a criminal complaint in relation to such a criminal offence, the opportunity to exercise their right to indemnity and the possibility to influence the course of criminal prosecution.

2 Research studies regarding the use of mobile devices among students and employees in organisations

This chapter presents two research studies focusing on the use of mobile devices that involved two different structures of respondents' populations. It also presents the results of both research studies, which provide a comprehensive overview of the use mobile devices in individual inter-generational populations. These two researches are the only two researches conducted in two different generational populations in close temporal proximity in Slovenia. They contained similar questions, therefore their results could be compared.

2.1 Methodology

The following sections present key demographic data obtained by the two research studies, which illustrate individual characteristics of the two population groups which participated in the studies.

In December 2011, a research study entitled Awareness of Threats to Mobile Devices was conducted among the student population with a view to obtain relevant information regarding their level of awareness and knowledge of threats that young users, particularly students, are exposed to when using mobile devices and establish whether they are familiar with the functioning and use of security protection. The research was carried out via an online questionnaire, which was available in December 2011 at the "1ka" portal (www.1ka.si) for a period of three weeks. Information about the research was sent to students via e-mail, online social networks and personal invitations. Collected data were then analysed by applying the SPSS tool. The analysis covered 281 questionnaires. Some respondents did not provide answers to all questions, which was duly taken into account by the researchers when analysing the sample. The analysis of responses thus contains information about the population that was actually represented in analysed samples. Most respondents were between 21 and 25 years of age, which is not surprising as the research was primarily conducted among students. There were 61.5 percent of female and 38.5 percent of male respondents. Most respondents completed secondary education.

The second research study was conducted between May 2012 and February 2013 via the same portal (www.1ka.si). The purpose of the research was to obtain a comprehensive overview of the actual state-of-play related to the use of mobile devices among employees of individual organisations. During the aforementioned period, more than 600 users of mobile devices from 34 different organisations in Slovenia responded to the online questionnaire. Almost 50 percent of all questionnaires were incomplete, which is why they were excluded from further analysis. The remaining part of this paragraph presents results obtained through questionnaires that reveal certain demographic data of respondents participating in the research. Most respondents completed post-secondary, third level or university education (65 percent), 19 percent completed secondary education, while 15 percent held a master's or doctoral degree. On the basis of these results, it is possible to claim that respondents have a level of education that provides them with a sufficient degree of sophistication and general knowledge that enable them to recognise threats to mobile devices, the severity and value of consequences arising from the actualisation of such threats and the importance of using security protection. Respondents were also asked about the size of organisations in which they are employed. The largest share of respondents, i.e. 81 percent, works in large enterprises, while the shares of respondents employed by other companies are substantially lower. Two percent of respondents work in micro-enterprises, eight percent are employed in small-sized enterprises and nine percent work in medium-sized enterprises. The ratio between the size of the organisation and the number of persons responding to an online survey is understandable. A higher frequency of responses and a larger size of respondents' shares are, in fact, directly correlated to the size of an organisation. The larger the company, the higher the number of employees that respond to the survey.

Employees of organisations participating in the research were also asked about their area of work. 300 respondents answered this question. Such data are extremely important as they reveal the skills and competences employees must have in order to work in a specific field and show whether they use mobile devices appropriately and whether they comply with information security requirements. Most respondents (46 percent) work in the field of information technology. Fifteen percent of respondents work in the field of economy, fourteen percent in the field of management and an equal share in marketing. Eleven percent of respondents work in the field of security. There were no respondents from the field of education and training.

Demographic data indicate some basic characteristics of respondents who participated in the two aforementioned research studies. The following section presents the results of analyses of individual questions, which are similar in both research studies in terms of their substance and serve as a basis for establishing differences between individual responses and draw conclusions with respect to future development trends.

2.2 Research results

In order to be able to use appropriate protection against threats that all users of mobile devices are exposed to, they must, first and foremost, be familiar with and aware of such threats. This is why students were asked whether they were familiar with the following threats.

The results to this question represent the degree to which respondents are familiar with individual threats that users of mobile devices are exposed to. Theft is the most widely recognised threat (89.4 percent), followed by viruses (83.1 percent). This is not surprising since such threats have existed for quite some time. On the other hand, the fact that respondents are not well acquainted with contemporary, advanced and rapidly growing threats, such as rootkit (14 percent) and malware infections (28.9 percent), is somewhat alarming. A similar question was posed to users employed in different organisations. Their results show the extent to which employees are aware of the consequences of threats or the possibilities for the realisation of individual threats to mobile devices. For each of the aforementioned threats, respondents were asked to choose between four possible answers (1 – I am not familiar with the threat; 2 – I do not believe this could happen to me; 3 – I believe this could happen to me; 4 – This has already happened to me). The largest share of respondents stated that they believed every single threat listed in the questionnaire could happen to them while using a mobile device. This means that they are aware of the danger, however, they mostly fail to use appropriate security solutions for protecting themselves from such threats, as subsequently demonstrated.

In order to protect one's mobile device against various threats, one needs to be familiar with some of the most basic security measures. According to the obtained responses from students, PIN numbers for unlocking SIM cards are the most frequently used security solution, which was to be expected, since providers of mobile telephony services normally build such a measure into the SIM card itself. On the other hand, the fact that large numbers of respondents are familiar with the possibility of using a PIN number for accessing individual applications on their mobile device, but are not using it (56.8 percent), is rather alarming, as this is automatically enabled by all mobile devices. Remote device wipe is also one of the proposed solutions. It may be used in case of loss or theft of a mobile device. However, only 40 percent of respondents are familiar with this security solution and still decide not to use it, while 52 percent of respondents are not acquainted with it at all. Given the fact that data from question regarding knowing threats clearly show that the majority of respondents believe that theft is a viable threat, the knowledge and use of this security solution would, in fact, be more than beneficial. Respondents should also gain knowledge regarding all of the aforementioned aspects of mobile device protection during their education and training process. Such education and training may generally refer to all potential solutions for protecting different models of mobile devices or cover specific types of mobile devices and particular software solutions. The second research study involving employees in organisations also looked into the use of security solutions available for mobile devices.

When analysing responses to this question, the "Yes" and "Yes, occasionally" answers were reasonably grouped into a common YES group, while "No, but I could" and "No, as I am not familiar with it" responses were grouped under NO. The authors of this paper were interested in determining whether individual variables bear any statistically significant differences between the YES and NO groups of responses, i.e. between those respondents who use security protection for their mobile devices and those who do not.

Since these two variables are dichotomous (they only have two values), a binomial test was used (to compare the two groups with each other), which is carried out in two steps. If the p-value is lower than 0.05, the difference related to individual variables is proven to be statistically significant in favour of the group having a larger share. If, however, the p-value is higher than 0.05, it is deemed that the two groups are equal and that there are no differences between them. A statistically significant difference for nearly all variables was observed between the group of respondents using security solutions and the one that does not. Such a difference is absent only in the following two variables: providing a safe and secure connection between mobile devices and organisations' information systems (e.g. VPN connection) and knowledge gained during education and training courses focussing on the functions and security of mobile devices. In variables showing statistically significant differences, however, the share of almost all variables (with the exception of the following: passwords or PIN numbers for accessing a mobile device, synchronisation of mobile devices' content) is tipped in favour of the group not using such a solution. This means that the share of those who do not use the aforementioned security solutions is prevailing.

Results of both research studies demonstrate the level of knowledge and awareness of threats to mobile device users and the extent to which they use security solutions. In order to evaluate risks, one must also obtain information regarding data that users themselves download to or access via their mobile devices. The student population was asked about the types of data they keep on their mobile device. The largest share (98.5 percent) of students responded that they use mobile devices to save mobile phone numbers, followed by photographs and images (94.1 percent) and contact lists (67.8 percent). The aforementioned types of data seem perfectly reasonable given the target group of respondents, but they also represent personal data. The fact that 25.4 percent of respondents keep work-related e-mail addresses and certificates (2.4 percent) on their mobile devices is much more alarming. The group of respondents employed in organisations was also asked about the type of content they keep on their mobile devices. Respondents simultaneously selected several possible answers. The majority of respondents keeps data necessary to get in contact with other people on their mobile devices (93 percent of all 300 respondents chose this particular variable). If one considers the historical perspective, this was to be expected since the principal goal of mobile devices is to enable communication between individual users or their devices, which is why one requires contact details. 87 percent of respondents keep photographs and images on their mobile device. It is interesting to note that the difference between the volume of personal and work-related e-mail amounts to merely four percentage points (work-related e-mail accounts for 60 percent of all e-mail, while personal e-mail represents 56 percent of the total e-mail volume). The same holds true for the proportion of respondents using personal and business calendar features on their mobile devices (61 percent of respondents use the personal calendar, while 57 percent of them use business calendar). Data also indicate the percentage of users who keep "sensitive data" on their mobile devices, i.e. data that may, for example, be considered confidential or business secrets or are of particularly personal nature (passwords, PIN number, etc.). A whopping 19 percent of respondents keep their charge card's PIN number saved on their mobile device, five percent of respondents keep their digital certificate (for online banking, access to business systems, etc.), and five percent use their mobile device to record their alarm systems' codes, online banking passwords, etc. Four percent of respondents revealed that they saved passwords for accessing business secrets and systems on their mobile devices. Two percent of respondents keep confidential data on their mobile device, while the same share of respondents also uses their devices to store confidential documents. These results clearly demonstrate that respondents use mobile devices to handle data that are of key importance both for their work and for the performance of the entire organisation. If such data were lost, both individuals and organisations would suffer significant damage.

3 Status of victims of security incidents

The realisation of a threat to the information security of mobile devices may represent a violation of the (personal) rights of users of such mobile devices. This is why the law must provide appropriate solutions that allow victims of security incidents to protect their rights. The following sections present different possibilities granted by the Slovene civil and criminal law that users of mobile devices may rely on when they fall victims of security incidents. Injured parties have several legal remedies at their disposal for protecting their personal rights. These are not mutually exclusive and failure to select a certain legal remedy does not normally affect the degree of validity of another legal remedy [9, 10].

3.1 Request to cease the infringement of personal rights

When a security incident occurs, it becomes imperative to prevent any further violation of rights. Users of mobile devices may do so by lodging a request to cease the infringement of personal rights as stipulated by the Code of Obligations (2007) according to which everyone has the right to request the court or any other competent authority to order that actions infringing the inviolability of persons, their personal and family life or any other personal right be ceased, that such actions be prevented or that the consequences of such actions be eliminated [7].

In this respect, one could first raise the issue of defining personal rights. These include all constitutional rights and fundamental freedoms [11]. Chapter II of the Constitution of the Republic of Slovenia (2016) thus regulates human rights and fundamental freedoms, including the right to personal dignity and safety, the general protection of the rights to privacy and personal rights and the specific protection of personal data [12]. The Slovene case law also refers to, for example, the right to reverence, i.e. memory of the deceased, and the rights to personal identity [13], the right to reputation and good name [14, 15], the right to privacy [16], etc. [17, 18, 19].

An infringement of personal rights results from any unjustified encroachment upon the aforementioned constitutional rights [11]. Such an infringement represents the only grounds on the basis of which legal protection according to Article 134 of the Code of Obligations (2007) may be requested. Furthermore, this possibility is only granted to persons who were directly injured, i.e. to victims whose rights were actually violated [11]. In principle, each violation of personal rights is unlawful, which stems from the absolute legal nature of personal rights as such [13]. However, an encroachment upon personal rights would be considered justified in cases listed in the Slovene Constitution (2016) or its legal order. In line with the general principles of the law of damages, this also includes the injured party’s consent to encroachment [16].

The Code of Obligations thus provides for three groups of claims. The first refers to the request to cease (prevent) all acts infringing someone’s personal rights provided that the infringement is still ongoing (preventive injunction) [11]. Upon receiving the injured party’s proposal, the court may order the infringer to pay a certain monetary sum to the injured party if they do not cease all acts infringing the injured party’s rights [17]. In practice, such a claim represents a request to remove certain publications, withdraw printed copies from circulation, remove photographs from internet portals, etc.

The second possibility refers to the prohibition of any subsequent actions infringing someone’s personal rights. This results in a prohibitory injunction being issued to the infringer in order to prohibit them from conducting any actions that would infringe the injured party’s rights in the future [11]. A previously proven infringement of personal rights is sufficient to demonstrate the existence of the risk of reoffending [16], if the threat of encroaching upon a certain personal right or the possibility of repeating the infringement is sufficiently specific [20]. The third possibility refers to the request to obtain redress for the violation of personal rights which may be achieved by any action or conduct allowing the attainment of redress, including the publication of a judgement or a correction [11]. In case of an infringement of personal rights, a court may order the publication of a judgement or a correction at the injurer's expense or order the injurer to retract a statement by which the infringement was committed or do anything else through which it is possible to achieve the purpose achieved via compensation [7, 19].

Injured parties or users of mobile devices also have the possibility of enforcing their claims in line with the Enforcement and Securing of Civil Claims Act (2015). Such claims refer to non-pecuniary claims or obligations to perform actions that can only be performed by debtors or obligations to cease or omit such actions. In this case, a court sets an appropriate deadline in which a debtor must comply with certain obligations. In addition, the court also imposes a fine in case the debtor does not meet their obligations within the set deadline. The court may impose a maximum fine of 10,000 EUR for a natural person, and a maximum fine of 500,000 EUR for legal entities and entrepreneurs. If the debtor does not comply with the obligations, the court conducts an ex officio enforcement on the basis of a decision concerning the imposition of a fine. The court subsequently issues a new decision in order to set a new deadline for the fulfilment of obligations by the debtor and imposes a new fine, which is higher than the one defined in the previous decision, in case the debtor would, once again, not comply with the obligations within the set deadline until the total sum of fines arising from individual decisions reaches a ten-fold amount of original fine [21].

3.2 Interim orders according to the Enforcement and Securing of Civil Claims Act (2015)

Interim orders are also relevant, as victims of security incidents are interested in ensuring that the courts prohibit any further violations of their personal rights as soon as possible and temporarily, i.e. for the duration of civil proceedings in which victims exercise the protection of their rights. This possibility is also provided in the Enforcement and Securing of Civil Claims Act (2015) according to which an interim order may be issued prior to the initiation of proceedings before a court, during the course of proceedings, as well as after the completion of such proceedings provided that the conditions for enforcement remain present [21]. Naturally, it is in the victims' interest to request that such an interim order be issued as soon as possible following the detection of security incidents and the violation of their rights.

Courts issue *interim orders to secure non-pecuniary claims* if victims of security incidents are able to demonstrate the likelihood of the existence of a claim or that a claim against the debtor will arise. Victims must also demonstrate the likely risk that the enforcement of the claim will be impossible or rendered considerably more difficult; that the order is necessary in order to prevent the use of force or avert the occurrence of irreparable damage; or that the debtor will not suffer more detrimental consequences than the creditor if the interim order issued is proved to be unfounded in the course of proceedings. The risk of the occurrence of irreparable damage is particularly relevant when personal rights are violated as a result of a security incident. It is normally deemed that a risk was demonstrated if the claim is to be enforced abroad, which is often the case in security incidents, as communication via mobile devices usually involves an international element (e.g. as a result of cloud computing). In order to secure non-pecuniary claims, courts may issue any interim order enabling the achievement of the purpose of such protection or determine an appropriate security [21].

3.3 Status of victims of security incidents in criminal proceedings

The realisation of threats to mobile devices' information security may also represent a criminal offence, such as the abuse of personal data, the violation of secrecy of the means of communication, the violation of moral or material copyright, attacks on information systems, etc [8]. In Slovene criminal law, an injured party is defined as a person whose personal or property rights have been violated or jeopardised as a result of a criminal offence [22].

The first issue that is raised in this respect is whether and when does the injured party have the possibility or duty to report a criminal offence that was committed against them, since this is not always in their interest. The Criminal Procedure Act (2014) contains a general arrangement regarding criminal complaints, which is also relevant for injured parties. Any person, including the injured party, may thus report a criminal offence which is liable to an ex officio public prosecution, while the relevant legislation defines cases where the failure to report a crime is considered a criminal offence itself. The Criminal Procedure Act (2014) stipulates that all state agencies and organisations having public authority are bound to report criminal offences liable to public prosecution [22]. In addition, the Criminal Code-1 (2017), which defines the failure to provide information of a crime or its perpetrator as a criminal offence, also stipulates that whoever knows of a criminal offence for which the sentence of no less than fifteen years of imprisonment or life imprisonment is prescribed has the obligation to report such a crime; the same reporting obligation is extended to officials who, during the performance of their official duties, come to know of a criminal offence for which the punishment of more than three years of imprisonment is prescribed under the statute [8]. In line with the described arrangement, an injured party who does not have the status of an official is not obliged to – but may – report a security incident that they have fallen victim to, since penalties for criminal offences usually committed in the scope of security incidents are shorter than fifteen years.

Injured parties also have the right to request the reparation of damage caused by a criminal offence either through an indemnification claim within criminal proceedings or by an action in civil litigation proceedings. In both cases, the reparation of damage incurred as a result of encroachment upon personal rights may be requested by the person entitled to assert such a claim in a civil action [22]. Given the fact that Slovene criminal courts are not in favour of deciding upon indemnification claims, particularly in cases related to the proceedings for the compensation of non-pecuniary damage, the possibility providing direct redress, i.e. a civil action, is considered more appropriate.

However, the mere infringement of personal rights does not mean that the plaintiff's claim for pecuniary compensation or request for the publication of correction will be successfully resolved. Such an infringement must be unlawful and must cause legally recognised damage to the plaintiff [23]. Even when dealing with an infringement of personal rights, one must be able to prove the existence of all general preconditions governing the fault-based liability for damages, unlawful conduct, damage, as well as the causal relationship between such conduct on one hand and the incurred damage and fault on the other [11, 18]. It is thus possible to assert both pecuniary and non-pecuniary damages. With respect to the latter, it is particularly important to consider compensation resulting from inflicting mental distress or fear on victims of security incidents [7, 19].

Another issue, i.e. do victims of security incidents have the possibility of influencing the course of criminal proceedings, is also relevant. Apart from the usually present desire for retribution, the relevant legal interests often encourage injured parties to institute criminal proceedings (as soon as possible), for instance with a view to facilitate the assertion of the liability for damages. However, injured parties have a direct possibility to influence the course of criminal proceedings against an alleged perpetrator only when criminal offences are prosecuted in a private action (e.g. in case of criminal offences against honour and reputation which are, in fact, relevant for this paper) [8]. With respect to criminal offences that are prosecuted ex officio, injured parties only have the possibility of influencing the course of criminal proceedings indirectly, for example in the event of summary proceedings; if the criminal complaint is filed by the injured party and the public prosecutor fails to submit a summary charge sheet and notify the injured party that they have dismissed the charge or adjourned criminal proceedings within a period of one month after receiving the criminal complaint by the injured party, the injured party is entitled to assume prosecution subject to filing the summary charge sheet with the court [22]. However, such a solution does not exist in regular criminal proceedings, thus failing to provide injured parties with a mechanism that would enable them to “force” a decision by the public prosecutor regarding the fate of a criminal complaint. Injured parties may influence the course of proceedings only if a settlement is foreseen or if the public prosecutor suspended criminal prosecution with the consent of the injured party [22] and if a subsidiary prosecution was initiated.

Criminal proceedings may also have negative consequences for victims of security incidents. Apart from secondary victimisation, injured parties must also bear high costs of criminal proceedings. Fees and necessary expenses for the private prosecutor, the injured party and the injured party acting as prosecutor, as well as their representatives and attorneys are considered as costs of criminal proceedings, but the represented party, i.e. the injured party, must pay them in advance, except where such fees and necessary expenses are to be charged to the state budget. Courts decide on the costs of proceedings upon their completion. Private prosecutors and injured parties acting as prosecutors are bound to repay the costs of criminal proceedings if the proceedings terminate in a judgement of acquittal, except where the ruling discontinuing proceedings or a judgement rejecting the charges have been passed through no fault of the private prosecutor or injured party acting as prosecutor[22]. This means that criminal proceedings may represent a substantial financial burden for the injured party, which may also affect their decision regarding the initiation and/or conduct of criminal proceedings or protection of their rights by means of such proceedings.

4 Conclusions

Data obtained by the two research studies presented in this paper demonstrate that the awareness of threats arising from the use of mobile devices is higher among the employees. However, similarly to the students' population, their use of security solutions is not adapted to contemporary sophisticated threats to mobile devices. The risk of losing data is thus even greater in both groups of respondents, since it is clear (this was demonstrated by both research studies indicating the use of data and services on mobile devices) that the boundary between personal and private use of mobile devices has truly and completely disappeared. This is much more understandable in case of employees, while it is rather surprising and alarming with respect to students, since they will soon become employees of various organisations and (increasingly) use mobile devices to access and download important data. How are they supposed to do that successfully when they are not aware of threats (which clearly shows their lack of knowledge with respect to the use of mobile devices and their protection), do not use appropriate security solutions and use mobile devices to access both personal and business data? Research shows that the lack of knowledge and rules concerning the use of mobile devices both in organisations, as well as in the sphere of education at the national and transnational levels, represents the main problem faced by today's population when using mobile devices and providing information security.

The overview of the Slovene legal system shows that users of mobile devices who fall victims of threats to mobile devices' information security have certain possibilities at their disposal to protect their interests, particularly to cease existing encroachment upon their personal rights and to prevent any future infringements both on the basis of civil and criminal law. The problem that occurs when implementing such legally awarded possibilities in practice stems from the gap between the need to react quickly in order to protect one's personal rights through civil and criminal proceedings, which are often lengthy, and the nature of mobile devices and contemporary electronic communications in general, which enable an instant dissemination of information and often include an international dimension. The latter also raises the questions of jurisdiction and applicable law in legal proceedings, which normally makes proceedings even more complex and exacerbates the injured party's position. This is why it is imperative for the victims of security incidents affecting mobile devices to be able to rely on a stable system ensuring the protection of their personal rights.

References

[1] International Data Corporation [IDC]. (2017). IDC – Press Release. Gathered from: https://www.idc.com/getdoc.jsp?containerId=prUS42628117

[2] Tiongson, J. (2015). Mobile app marketing insights: How consumers really find and use your apps. Think with Google. Gathered from: https://www.thinkwithgoogle.com/consumer-insights/mobile-app-marketing-insights/

[3] McAfee. (2014). McAfee Labs threats report. Gathered from: https://www.mcafee.com/us/resources/reports/rp-quarterly-threat-q1-2014.pdf

[4] McAfee. (2015). McAfee Labs threats report. Gathered from: https://www.mcafee.com/us/resources/reports/rp-quarterly-threat-q1-2015.pdf

[5] McAfee. (2016). McAfee Labs threats report. Gathered from: https://www.mcafee.com/hk/resources/reports/rp-quarterly-threats-dec-2016.pdf

[6] McAfee. (2017). McAfee Labs threats report. Gathered from: https://www.mcafee.com/us/resources/reports/rp-quarterly-threats-mar-2017.pdf

[7] Code of Obligations, Official Gazette of the Republic of Slovenia, No. 97/07 – official consolidated version.

[8] Criminal Code-1, Official Gazette of the Republic of Slovenia, No. 50/12 – official consolidated version, 54/15, 38/16 and 27/17.

[9] Šinkovec, Janez; Tratar, Boštjan, Obligacijski zakonik s komentarjem in sodno prakso [Code of Obligations with Commentary and Case Law]. Lesce: Oziris, 2001.

[10] Supreme Court of the Republic of Slovenia, II Ips 582/96, 1998.

[11] Plavšak, Nina; Juhart, Miha; Vrenčur, Renato, Obligacijsko pravo, splošni del [Law of Obligations, General Part]. Ljubljana: GV,

[12] Constitution of the Republic of Slovenia, Official Gazette of the Republic of Slovenia, Nos. 33/91-I, 42/97, 66/00, 24/03, 47, 68, 69/04, 69/04, 69/04, 68/06, 47/13, 47/13 and 75/16

[13] High Court of Ljubljana, II Cp 764/2009, 2009

[14] Supreme Court of the Republic of Slovenia, II Ips 340/2011, 2014.

[15] Lampe, Rok, Pravo človekovih pravic [Human Rights Law]. Ljubljana: Uradni list, 2010.

[16] High Court of Ljubljana, II Cp 2955/2013, 2014.

[17] Šinkovec, Janez, Pravice in svoboščine [Rights and Freedoms]. Ljubljana: Uradni list, 1997.

[18] Polajnar-Pavčnik, Ada, Temeljne pravice kot osebnostne pravice [May Fundamental Rights Be Considered as Personal Rights?], in: Pavčnik, Marijan; Polajnar-Pavčnik Ada, Wedam-Lukić Dragica (Eds.), Temeljne pravice [Fundamental Rights]. Ljubljana: Cankarjeva založba, 1997.

[19] Lampe, Rok, Sistem pravice do zasebnosti [The System for Exercising the Right to Privacy]. Ljubljana: Bonex, 2004.

[20] High Court of Ljubljana, I Cp 1308/2010, 2010.

[21] Enforcement and Securing of Civil Claims Act, Official Gazette of the Republic of Slovenia, Nos. 3/07 – official consolidated version, 93/07, 37/08, 45/08, 28/09, 51/10, 26/11, 17/13, 45/14, 53/14, 58/14 and 54/15.

[22] Criminal Procedure Act, Official Gazette of the Republic of Slovenia, No. 32/12 – official consolidated version, 47/13 and 87/14.

[23] High Court of Ljubljana, I Cp 106/2014, 2014.

Combat Mobile Evasive Malware via Deep Learning

IRFAN BULUT, ALI GOKHAN YAVUZ & RÜSTEM CAN AYGUN

Abstract Malware has become more harmful than in the past as the number of intelligent systems increased dramatically. Especially mobile device have become the predominant means of accessing personalized services such as email, banking, etc. However, this rapid deployment and extensive availability of mobile applications has made them attractive targets for various malware. Therefore one of the most important issues in cyber security has become the detection of previously unknown malware in the shortest time possible in order to stop it from becoming common hazards and to harm users. For this purpose, machine learning methods were applied to detect and classify malware. But methods are commonly based on information gathered via dynamic analysis to achieve better accuracy with machine learning based techniques. So it is necessary to execute the unknown software to collect features which can't be obtained via static analysis. But malware authors usually bypass this automated malware analysis process and consequently true detection rates fall.

Consequently in this study, we present a novel model based on deep learning for the prediction of mobile malware without requiring execution in an isolated environment. After optimizing their weights with automatic encoder, we obtained an accuracy of 93.67% with a subsequent multilayer perceptron.

Keywords: • Android • malware • malicious software • evasive malware • anti-analysis • anti-virtualization • deep learning • information security • cybersecurity •

CORRESPONDENCE ADDRESS: Irfan Bulut, Yildiz Technical University, Computer Engineering Department, Mah. Davutpaşa Caddesi 34220 Esenler- İstanbul, Turkey, e-mail: msc.irfanbulut@gmail.com. Ali Gokhan Yavuz, Ph.D., Assistant Professor, Yildiz Technical University, Computer Engineering Department, Mah. Davutpaşa Caddesi 34220 Esenler- İstanbul, Turkey, e-mail: gokhan@ce.yildiz.edu.tr. Rüstem Can Aygun, Research Assistant, Yildiz Technical University, Computer Engineering Department, Mah. Davutpaşa Caddesi 34220 Esenler- İstanbul, Turkey, e-mail: can@ce.yildiz.edu.tr.

DOI https://doi.org/10.18690/978-961-286-114-8.7 ISBN 978-961-286-114-8

Available at: http://press.um.si.

1 Introduction

Malware (***Mal**icious soft**ware***) is one of the most effective arsenal used by the cyber criminals [1]. Malware cause disruption to the operation of infected systems, convert them into slaves to be used in botnets, compromise the systems' and/or users' security or access sensitive information, make premium calls and/or send SMS advertisement spams on mobile systems without the users' permission and/or information.

In detecting and combatting malware, one of the most common methods is the use of the legacy Antivirus (AV) systems. These AV systems basically rely on a very simple signature-based method. The signature can either be a checksum generated via a MD5/SHA based hash function applied to known malware, or a specific string representing a URL or the name of an imported function and the like. Consequently the signature extracted from the unknown program is compared to a database of signatures to detect malware. In response to the signature based detection methods, malware authors increasingly exploited techniques such as changing the order of the code or packages within the code, adding unnecessary code or encrypting the code itself, each of which was effective enough to defeat the effectiveness of the signature based detection systems. Thus, we can safely say that the legacy AV systems could not be considered as a defense mechanism against today's modern malware.

Although the first malware appeared on desktop systems, today mobile systems are the most attractive targets of malware authors as in recent years, mobile systems became the dominant technology for performing most of our daily activities. According to IDC Quarterly Mobile Phone Tracker, Android would continue to capture roughly 85% (2017 Q1) of the worldwide smartphone volume and each day more and more mobile applications are being developed with user convenience at the focus. Unfortunately, user convenience usually means less security. Thus mobile malware authors especially target application markets with low barriers to enter, such as Android application markets. As a result the growing amount and diversity of Android malware combined with the significantly weakened effectiveness of the conventional defense mechanisms, researchers were forced to study more effective ways of detecting and isolating modern malware, which in turn led to isolated dynamic analysis with artificial neural networks and/or machine learning algorithms.

1.1 Aim and Scope

Isolated systems, i.e. sandboxes, are the most promising as well as the weakest chain in modern malware detection. Malware authors integrate code into modern malware to find out if it is being analyzed in an isolated system for detection, thus it is of utmost importance for the isolated systems to hide their presence. Unfortunately, no perfect sandbox system exists and consequently malware authors could easily trick such sandboxes and make them classify malware as benign programs. Therefore, in this paper we propose a solution which does not require the execution of the malware in a sandbox in order to detect its true nature.

1.2 Contribution

The proposed method is based on deep learning techniques and it is aimed at detecting Android based malware. We use Android applications' permissions which are requested from the user either during the installation or during the first action that requires the corresponding permission. These permissions are used as features to classify the application and they can be obtained without executing the application. During our tests,

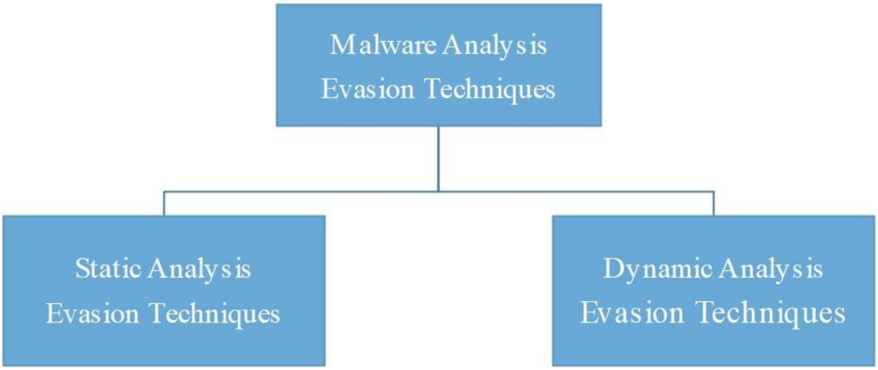

Malware Analysis Evasion Techniques
Static Analysis Evasion Techniques
Dynamic Analysis Evasion Techniques

Figure 1: Malware analysis evasion techniques

3229 benign and 1668 malicious applications were used to test the accuracy of our proposed deep learning model. The results show that the proposed model could identify malware with 93.67% accuracy.

1.3 Outline

The remainder of this paper is organized as follows: Section II discusses malware analysis evasion techniques. Section III gives information about related works. Section IV discusses the characteristics of the dataset and the feature extraction process. Section V describes our proposed deep learning model. In Section VI, the experimental results are discussed and analyzed. And, finally Section VII concludes the work and describes future directions.

2 Malware analysis evasion techniques

Malware implements the self-protection by evasion of analysis techniques [2][3][4][5]. As shown in Figure 1, evasion of analysis strategy is divided into two basic categories: static analysis evasion techniques and dynamic analysis evasion techniques. The static analysis evasion techniques use the method of packing and code obfuscation to disturb disassembly and identification of control flow, whereas dynamic analysis evasion techniques detect operating system environment to realize the anti-tracking for the debugger and sandbox systems.

2.1 Static Analysis Evasion Techniques

Malware authors use several techniques to protect themselves against static analysis and confound the analysis process [2]. The most common techniques are as follows:

Packer: Early malware only uses encryption to evade detection and analysis. The encrypted malware is mainly composed of two parts: the decryption part and the encrypted code body. When such a malware is about to execute, the two parts are loaded into the memory simultaneously. The execution begins with the decryption part in order to decrypt the encrypted code body, and then control is passed to the unencrypted version of the code to perform the actually intended function. Malware components realizing compression and encryption together are referred to as packers.

Code Obfuscation: Obfuscation is a technique that makes binary and textual data unreadable and/or hard for not being reverse engineered to understand. Some obfuscation examples used in a lot of malware are exclusive or operation, *Base64 encoding, runtime packers, exclusive or operation, ROT13* etc. Code obfuscation technology is getting more and more attention acting as a complementary security branch of code/data encryption technology. Code obfuscation is widely used not only in the field of software protection but also used by malware for the protection of malware to evade detection and analysis.

Information Hiding: This method modifies *strings, variables, debugging information, import table,* and other information within the executable. The purpose is hiding information which will be helpful to the analyst. Particularly, an experienced analyst could roughly guess the behavior of a given program by examining its import table. Therefore, modifying the import table is very important for malware programs to evade static analysis techniques.

2.2 Dynamic Analysis Evasion Techniques

Evasion of dynamic analysis primarily depends on the detection of the current operating environment. To this end, the malware tries to obtain as much information as possible from the runtime environment to correctly fingerprint any given analysis tool. Thus, this process includes detection of *debuggers, virtual machines, and monitoring tools* [4]. The current state of art techniques are:

- **Anti-debugging methods:** Anti-debugging methods refer to techniques that prevent the analysis of some parts of binary files in debugging environments by executing different execution flows, or by simply exiting the execution. Various anti-debugging techniques exist such as *API based anti-debugging, hardware based anti-debugging, timing based anti-debugging, detection of handles, services,* etc.
- **Detection of monitoring tools:** In the process of analyzing malware, many monitoring tools such as *Procmon, Process Explorer, Regshot, Netcat, Wireshark, INetSim,* etc. are used. If malware detects such a monitoring tool is

either installed or running in the system it suppresses the execution of malicious parts and carries out its normally intended behaviour.

- **Detection of sandbox:** In addition to the debugger, malware authors focused on sandbox detection methods. In fact, many different mechanisms have been published regarding sandbox detection [6][7][8]. The details of sandbox detection mechanisms will be explained in more detail in the next section.

3 Related work

Android platform provides several official security solutions that harden the installation and execution of malware, such as the Android permission system and Google Bouncer. Google Bouncer is a service provided by Google Play, the official Android market, since 2012, which aims at automatically scanning applications (*both new and previously uploaded ones*) and developer accounts in Google Play with its reputation engine and cloud infrastructure. Even if Bouncer adds another line of defense to the Android security, it still has many limitations. First, Bouncer can only scan Android applications for limited time, which means a malware can easily bypass it by not doing anything malicious during the scan phase. Second, no malicious code needs to be included in the initial installer when it gets scanned by Bouncer. In this case, the malware can have an even higher probability to evade Bouncer's detection. Once the application passes Bouncer's security scan and gets installed on a real user's Android device, then the malware can either download additional malicious code to run or to connect to its remote Command-and-Control (C&C) server to leak data or to receive further commands. Similar vulnerabilities are present for all automatic malware detection systems. A lot of studies have been done to detect malware for different platforms (*personal computers, servers, mobile devices, and Internet of Things (IOT) devices*). It is very difficult to successfully generate signatures which can be used to prevent new attacks, so the conventional methods are usually ineffective against zero-day malware [9][10]. Several approaches have been suggested to improve the signature generation process. Here we briefly review several of them.

- In [11], using support vector machines (SVM), J48 Decision Tree algorithms, and bagging machine learning methods on a data set consisting of permissions and API (*Application Programming Interface*) calls, achieved accuracy was between 92.36% and 96.88%.
- In [12], authors used Bayesian, Classification and Regression Tree (CART), J48 Decision Tree, Random Forest, and Sequential Minimal Optimization (SMO) methods to detect malware using application permissions only. The reported accuracy was between 72.78% and 94.90% depending on the classifier.
- In [13], authors used automatic encoder-based deep ANN method, SVM, Decision Tree (DT), ANN, Naive Bayes (NB) methods to detect malware using Android kernel system calls. The reported accuracy was between 77.94% and 93.68% depending on the classifier. It was shown that a deep ANN classifier outperformed various machine learning classifiers.

- In the study [14], using Multi-Layer Perceptron (MLP), J48 Decision Tree, k-Nearest Neighbor (k-NN), Random Forest, and NB methods on a data set consisting of API calls, achieved highest accuracy was 83%.
- In another study, using deep learning, two separate data sets were created from the permissions that the applications requested at installation and at run time, and accuracies of 87.1% and 80.9% were obtained respectively using automatic encoders [15]. According to this study, the analysis made using the permissions requested at the time of installation ensures more accurate results.
- In [16], the proposed SVM based model takes into account both API calls and requested application permissions that were labeled as dangerous. Initial accuracy of 81% was increased to 86% after labeling certain API calls and application permissions as dangerous and retraining the model.

4 Dataset and feature extraction

By default, all Android applications work in an isolated environment. If they want to access camera, contacts, access, data, etc. outside the sandbox, they need the user's permission to do so. And users either do not pay enough attention what permission(s) they are granting to the application or they are not aware of the security implications of a particular permission. Thus applications do generally run with all the permission they asked for.

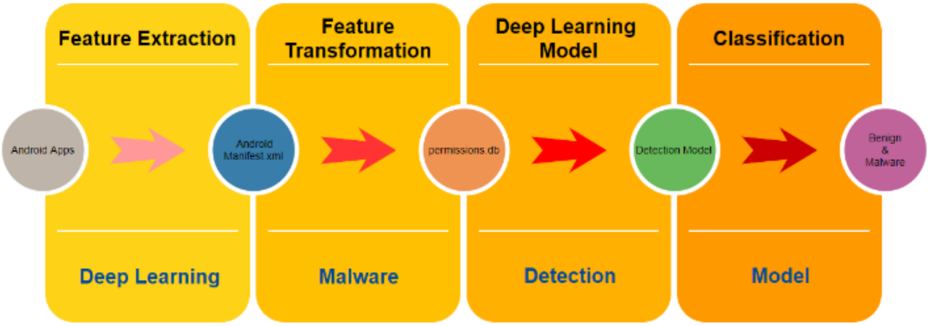


Figure 2: Feature extraction and deep learning malware detection model

Table 1: Dataset distribution

Data Class	Subdata Class		
	Training	**Test**	**Validation**
Benign	2261	484	484
Malware	1166	251	251

In this study, our proposed deep learning based malware detection model uses the permissions requested by an Android application to classify it either benign or malicious. These permissions are placed in the xml file named AndroidManifest.xml within the application's executable. These permissions are the most qualified attributes to understand all the possible activities that the application could do on the system and they can be obtained without executing the application.

In terms of static analysis, all the information we need is contained in the executable itself. As shown in Figure 2, after decompressing the apk file with the apktool, we mainly focus on parsing the AndroidManifest.xml. A total of 147 distinct permissions are defined for Android applications. These permissions are expressed in XML format. Depending on the requirements of the application, a varying number of permissions is contained in the AndroidManifest.xml. Thus, to request a specific permission its symbolic name must be present in the AndroidManifest.xml. For our dataset we had only 138 distinct permissions. We scanned all the AndroidManifest.xml files and eliminated permissions that were common to all of the applications. As a result,138 permissions were reduced to 128 permissions. The resulting 128 distinct permissions were used as features.

In the scope of the study, our dataset consisted of 3229 benign and 1668 malicious mobile applications. The applications for the benign class was chosen from applications that have the highest download rate in the Android market, and have been on the market for a long time without being tagged as malware. The malware samples were obtained from the Comodo security company. The data set used for the training, testing and verification phases of the system is divided into sub-data sets as shown in Table 1, with 70% training, 15% testing and 15% verification. Since all attributes in the data set are binary ("0" or "1), there is no need to perform any normalization.

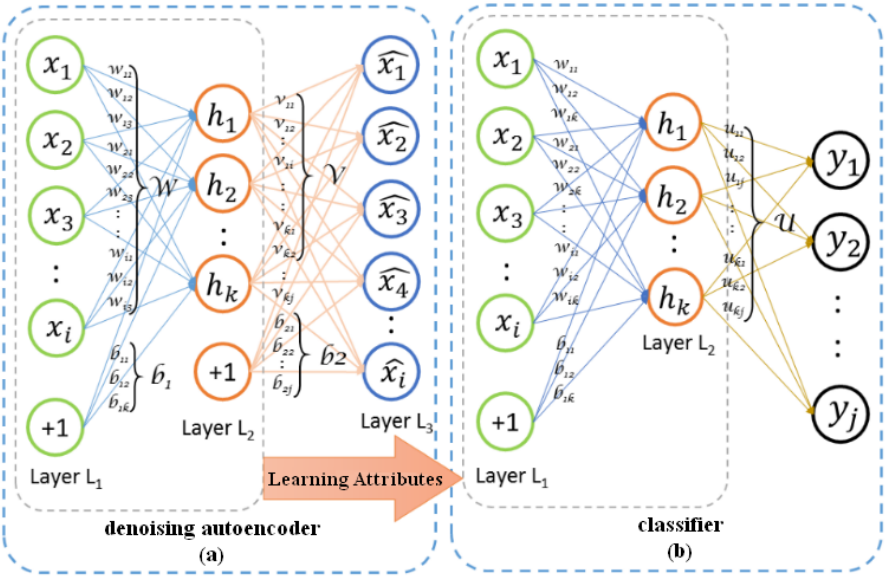


Figure 3: Deep neural networks training process (a) unsupervised learning attribute (b) supervised classifier training [18]

5 Deep learning model

Deep learning is a new area of machine learning research that imitates the human brain working mechanism and has gained increasing attention in the field of artificial intelligence. It has motivated a great number of successful applications in speech recognition, image classification, and natural language processing. Preliminary work in deep learning as it applies to Android malware detection [17]. In the proposed method, denoising autoencoder (DA) and multilayer perceptron (MLP) algorithms are used for malware detection as shown in Figure 3. Denoising autoencoders can be added subsequently to create a deep neural network model. DA based deep MLP consists of pre-training and fine tuning stages. In the first stage, the aim is to discover important features by using unlabeled data and then to use these discovered features to improve the performance of the supervised MLP classifier [19][20].

As shown in Figure 3 (a), the DA is trained with unlabelled data. At this stage, each hidden layer of deep neural network is considered as a separate autoencoder. In addition, these autoencoders are trained in a layer-by-layer manner starting from the shallowest layer to the deepest layer. This type of training is called greedy-wise training. The aim of the greedy-wise training is to ensure that the weights are tuned to the optimum level prior to the MLP supervised training. The error function of a MLP initialized with random weights could be stuck at the local minimum values while being trained by the gradient descent method and it reduces the model's performance.

The weights of the MLP are updated to make supervised training suitable thanks to the pre-training stage. After the pre-training stage, in the fine-tunning stage, the MLP with optimised weights is trained using labelled data to decrease the classification error as shown in Figure 3 (b) and the training continues until the weights of the whole model reach their ideal values. In this study, sigmoid function was selected as neuron activation function in DAs and softmax activation function was used at the output layer of the classifier. In addition, early-stopping and L1, L2 regularization (weight decay) methods were used to prevent the model from overfitting.

Table 2: Experimental results of machine learning algorithms

Algorithms	Precision	Recall	F1 Score	Accuracy
Naive Bayes	0,858	0,859	0,858	0,859
Logistic Regression	0,909	0,910	0,909	0,91
MLP	0,932	0,932	0,931	0,931
SVM	0,909	0,910	0,909	0,910
Bagging	0,919	0,919	0,919	0,918
J48	0,919	0,920	0,920	0,919

Table 3: Experimental results of our Deep Learning Model

Data Class	Result		
	Precision	Recall	F1 score
Benign	0.919	0.971	0.944
Malware	0.937	0.836	0.883
Weighted Average	0.925	0.925	0.923

6 Experimental results

In this study, the deep learning model is comprised of the input layer, three hidden layers, and an output layer to detect malware. The input layer of the model has 128 neurons because of the data set contains 128 distinct features. We have determined the neuron counts of the three hidden layers as 153 separately as a result of excessively parameter optimization searches in order to achieve optimal performance in pre-training stage. Finally, two neurons were used in the output layer be able to do two-class classification, *benign* and *malware*.

As shown in Table 3, the model performed with 91.9% and 93.7% precision value and 97.1% and 83.6% recall value for benign and malware classes, respectively. Given these results, the model shows a higher success in classifying benign data compared to malware data. It also describes the overall performance of the malicious software with the 92.3% F1-measure, which was found by the weighted average method for all benign and malware classes. In addition, the accuracy of the model on the test data is 93.67%.

We have also applied several other machine learning algorithms such as naive Bayes, logistic regression, MLP, SVM, Bagging, and J48 to obtain comparable results with our deep learning based proposed models. Weka was used to run these algorithms on our data set. As shown in Table 2, the closest result to the deep learning model was obtained with the MLP algorithm with a accuracy rate of 93.19%. The reason for the high success of MLP is that the number of samples in the dataset is limited. So, the MLP method yielded results close to the deep learning method without being affected by local minimum problem.

7 Conclusion

In this work we used 3229 benign and 1668 malicious Android applications to test the accuracy of our proposed deep learning model. The proposed deep learning model is based on a denoising autoencoder followed by a multi-layer perceptron. The test results show that our proposed model could identify malware with 93.67% accuracy. This accuracy is equivalent to the success rates achieved by machine learning methods using attributes that are obtained through dynamic analysis and is particularly superior in terms of not being affected by analysis evasion techniques.

References

[1] R. S. Pirscoveanu, S. S. Hansen, T. M. T. Larsen, M. Stevanovic, J. M. Pedersen and A. Czech, “Analysis of Malware behavior: Type classification using machine learning," 2015 International Conference on Cyber Situational Awareness, Data Analytics and Assessment (CyberSA), London, 2015, pp. 1-7.

[2] Moser, C. Kruegel and E. Kirda, “Limits of Static Analysis for Malware Detection," Twenty-Third Annual Computer Security Applications Conference (ACSAC 2007), Miami Beach, FL, 2007, pp. 421-430.

[3] Xu Chen, J. Andersen, Z. M. Mao, M. Bailey and J. Nazario, “Towards an understanding of anti-virtualization and anti-debugging behavior in modern malware," 2008 IEEE International Conference on Dependable Systems and Networks With FTCS and DCC (DSN), Anchorage, AK, 2008, pp. 177-186.

[4] Chen, Xu, et al. “Towards an understanding of anti-virtualization and anti-debugging behavior in modern malware," Dependable Systems and Networks With FTCS and DCC, 2008. DSN 2008. IEEE International Conference on. IEEE, 2008

[5] Fukushima, Yoshiro, et al. “A behaviour based malware detection scheme for avoiding false positive," Secure Network Protocols (NPSec), 2010 6th IEEE Workshop on. IEEE, 2010

[6] Raffetseder, Thomas, Christopher Kruegel, and Engin Kirda. “Detecting system emulators. International Conference on Information Security," Springer Berlin Heidelberg, 2007

[7] Vidas, Timothy, and Nicolas Christin. “Evading android runtime analysis via sandbox detection," Proceedings of the 9th ACM symposium on Information, computer and communications security. ACM, 2014

[8] Boomgaarden, Jacob, et al. “Challenges in emulating sensor and resource-based state changes for Android malware detection," Signal Processing and Communication Systems (ICSPCS), 2015 9th International Conference on. IEEE, 2015

[9] E. Filiol and S. Josse. “A statistical model for undecidable viral detection," Journalin Computer Virology, 3(2):65-74, 2007

[10] Y. Tang and S. Chen. “Defending against Internet worms: A signature based approach," In Proceedings of IEEE INFO COM, pages 1384- 1394, Miami, Florida, 2005

[11] N. Peiravian and X. Zhu, “Machine Learning for Android Malware Detection Using Permission and API Calls," 2013 IEEE 25th International Conference on Tools with Artificial Intelligence, Herndon, VA, 2013, pp. 300-305.

[12] U. Pehlivan, N. Baltaci, C. Acarturk and N. Baykal, “The analysis of feature selection methods and classification algorithms in permission based Android malware detection," 2014 IEEE Symposium on Computational Intelligence in Cyber Security (CICS), Orlando, FL, 2014, pp. 1-8.

[13] S. Hou, A. Saas, L. Chen and Y. Ye, “Deep4MalDroid: A Deep Learning Framework for Android Malware Detection Based on Linux Kernel System Call Graphs," 2016 IEEE/WIC/ACM International Conference on Web Intelligence Workshops (WIW), Omaha, NE, USA, 2016, pp. 104-111.

[14] M. Z. Mas’ud, S. Sahib, M. F. Abdollah, S. R. Selamat and R. Yusof, “Analysisof features selection and machine learning classifier in android malware detection," 2014 International Conference on Information Science & Applications (ICISA), Seoul, 2014, pp. 1-5.

[15] Xu, L., Zhang, D., Jayasena, N., Cavazos, J. “Hadm: Hybrid analysis for detection of malware," SAI Intelligent Systems Conference 2016 September 21-22, London, UK, 2016

[16] W. Li, J. Geand, G. Dai, “Detecting Malware for Android Platform: AnSVM-Based Approach," 2015 IEEE 2nd International Conference on Cyber Security and Cloud Computing, New York, NY, 2015, pp. 464-469.

[17] Z. Yuan, Y. Lu, Z. Wang, and Y. Xue, “Droid-sec: Deep learning in Android malware detection," in Proceedings of the 2014 ACM Conference on Special Interest Group on Data Communication (SIGCOMM, poster), 2014, pp. 371-372.

[18] Javaid, A.; Niyaz, Q.; Sun, W.; Alam, M. “A Deep Learning Approach for Network Intrusion Detection System," In Proceedings of the 9th EAI International Conference on Bio-inspired Information and Communications Technologies (formerly BIONETICS), New York, NY, USA, 3-5 December 2015; pp. 21-26.

[19] Y. Bengio, P. Lamblin, D. Popovici and H. Larochelle, “Greedy Layer-Wise Training of Deep Networks," in Advances in Neural Information Processing Systems 19 (NIPS’06), pages 153-160, MIT Press 2007

[20] P. Vincent, H. Larochelle Y. Bengio and P. A. Manzagol, “Extracting and Composing Robust Features with Denoising Autoencoders," Proceedings of the Twenty-fifth International Conference on Machine Learning (ICML’08), pages 1096 - 1103, ACM, 2008

Cipher, the Random and the Ransom: A Survey on Current and Future Ransomware

ZIYA ALPER GENÇ, GABRIELE LENZINI & PETER Y.A. RYAN

Abstract Although conceptually not new, ransomware recently re-gained attraction in the cybersecurity community: notorious attacks in fact have caused serious damage, proving their disruptive effect. This is likely just the beginning of a new era. According to a recent intelligence report by Cybersecurity Ventures, the total cost due to ransomware attacks is predicted to exceed $5billion in2017. How can this disruptive threat can be contained? Current anti-ransomware solutions are effective only against existing threats, and the worst is yet to come. Cyber criminals will design and deploy more sophisticated strategies, overcoming current defenses and, as it commonly happens in security, defenders and attackers will embrace a competition that will never end. In this arm race, anticipating how current ransomware will evolve may help at least being prepared for some future damage.

In this paper, we describe existing techniques to mitigate ransomware and we discuss their limitations. Discussing how current ransomware could become even more disruptive and elusive is crucial to conceive more solid defense and systems that can mitigate zero-day ransomware, yielding higher security levels for information systems, including critical infrastructures such as intelligent transportation networks and health institutions.

Keywords: • ransomware threat • • ransomware mitigation • malware • cybersecurity • survey •

CORRESPONDENCE ADDRESS: Ziya Alper Genç, Ph.D. student, University of Luxembourg, Interdisciplinary Centre for Security, Reliability and Trust, Maison du Nombre, 6, Avenue de la Fonte, L-4364 Esch-sur-Alzette, Luxembourg, e-mail: ziya.genc@uni.lu. Gabriele Lenzini, Senior Research, Interdisciplinary Centre for Security, Reliability and Trust, 29, avenue JF Kennedy, L-1855 Luxembourg, Luxembourg, e-mail: gabriele.lenzini@uni.lu. Peter Y.A. Ryan, Professor, University of Luxembourg, Interdisciplinary Centre for Security, Reliability and Trust, Maison du Nombre, 6, Avenue de la Fonte, L-4364 Esch-sur-Alzette, Luxembourg, e-mail: peter.ryan@uni.lu.

DOI https://doi.org/10.18690/978-961-286-114-8.8 ISBN 978-961-286-114-8

Available at: http://press.um.si.

1 Introduction

When installed on a system, a ransomware encrypts files or blocks functionalities and when the job is done it asks for a ransom. The victim is left with the choice between paying up and regain access to the files and functionalities or never being able to use the system again. The ransom is usually paid in anonymous cryptocurrencies, like Bitcoin [35], leaving the ominous transactions untraceable by the authorities.

No risk of being seized together with low development effort have made ransomware a very popular weapon in the arsenal of cyber criminals. Kaspersky reports that in every40secondsa business is attacked by ransomware and that frequency is fourfold for individuals [26]. A ransomware is therefore a type of malware but the nature of ransomware attacks significantly differ from the ones of conventional malware in terms of the economical damage and the recoverability. When a system is infected by a crypto-ransomware variant, the victim's files are encrypted using strong cryptography [8] and the recovery may not be possible. Given this situation even the Federal Bureau of Investigation (FBI) reportedly advises victims to simply pay the ransom [30]. However paying the ransom may not guarantee to obtain the decryption keys and recover the files [33].

While Microsoft Windows platform continues to be the major target of ransomware threat [40], recently a Korean hosting firm have been hit by a Linux ransomware and had to pay $1 million [17]. Ransomware are therefore cross-platform, targeting indiscriminately private citizens and companies, and able to hit many countries at once indistinguishably. The infamous WannaCry ransomware recently attacked more than 200 000 computers in 150 countries [15].

Attacks does not seem to slow down in the near future. According to a recent intelligence report by Cybersecurity Ventures, the total cost due to ransomware attacks is predicted to exceed $5 billion in 2017 [44]. Ransomware threat is likely to have more consequences than just economical e.g., denial of services, downgrade in service quality, lost of social trust or lost of trust in Information and Communications Technology (ICT). Ransomware attacks may even cause problems of civil liability and culminate lawsuits against victim companies and institutions from customers e.g., by patients whose care are delayed or hindered. All these circumstances draw attention of information security community and there have been several proposals to mitigate ransomware.

In this paper we review current defense techniques for ransomware, discussing their strong and weak points. Then, we discuss what potential strategies could ransomware designer implement to bypass current countermeasures to continue causing damage for an extended period: we introduce original ransomware variants that employ rootkit techniques and white-box cryptography, and, inspired by the cybersecurity incidents occurred in real-world applications, we point out new possible ransomware targets and attack types.

2 Background

Ransomware aims to extort money through preventing accessto data or functionality on victim's system. Cryptographic ransomware accomplishes this goal by encrypting files using strong cryptography [8] while holding the description keys so that victims are forced to pay the ransom to obtain those keys and regain access to their files. Another variant, the locker ransomware, reaches this aim via taking control of the victim's system and denies functionality. In this case, user data are untouched but the infected system becomes unusable.

2.1 Defense Systems

Current Mitigations. Existing anti-ransomware applications, excluding the inefficient and ineffective practice to back-up and restore files, can be grouped in three main families. The first includes defences which monitor an application's activity in real time in search for patterns that justify blocking a potential ransomware from working (*behavioral analysis*). The second contains defences that create the conditions to nullify or reverse the effect of a ransomware (*key escrow strategies*). The last one includes defences that isolate the binary of applications and analyze their code in search for calls to cryptographic operations which would reveal at least in potential the presence of a malicious intention (*detection of cryptographic primitives*). In detail:

- *Behavioral analysis:* In this approach, ransomware defense systems examine the behavior of an application and its interactions with the environment, e.g., file system activity, network connections and modifications on operating system (OS) components. There are various proposals in the literature that uses behavioral analysis approach. One of them, UNVEIL [27] generates an artificial user environment and monitors desktop lockers, file access patterns and I/O data entropy. Another one, CRYPTODROP [37] observes file type changes and measures file modifications using similarity-preserving hash functions and Shannon Entropy to recognize ransomware. Moreover, SHIELDFS[13] monitors low-level file system activities and collects the following features: *folder listing, file read/write/rename, file type and write entropy*. A ransomware is detected by comparing these characteristics with that of benign applications. Unlike the previous two, SHIELDFS can recover the files which are already encrypted before detection, though this capability comes with a significant performance overhead.
- *Key escrow:* In this approach, cryptographic materials generated by ransomware on the victim's system are obtained and held in escrow to later use for recovery. For instance, PAYBREAK [28] is a key escrow based mitigation system and works by intercepting cryptographic Application Programming Interface (API), extracting passed parameter sand storing them in a secure key vault. In the case of infection, the defense system tries to decrypt the encrypted files using the stored keys and parameters. However, this approach can succeed only if the cryptographic functions employed by the ransomware are correctly recognized and the parameters passed to the APIs are logged. While this is feasible for built-

in cryptographic functions on the host system, ransomware that utilizes third-party libraries can bypass detection through *obfuscation* [28] as we will discuss in Section 3.2.

- *Detection of cryptographic primitives:* In this approach, binary programs are analyzed to identify cryptographi coperations in their executable codes. To this goal, [19] traces the execution of applications and monitors I/O relationship in the program flow. Based on the occurrences of *bitwise arithmetic instructions* and *loops*, and relationships between the inputs and outputs of the program routines, heuristics are applied to recognize the cryptographic algorithms. On the other hand, [31] uses static analysis and Data Flow Graph (DFG) isomorphisms to identify cryptographic algorithms in the binary programs. Basically, this technique work as follows: First, the DFG of binary program is build. Next, the DFG in hand is normalized using rewrite rules in order to remove the variations due to complier optimizations. Finally, subgraphs which are isomorphic to graph signatures of cryptographic algorithms are searched in the DFG. A match directly flags that the corresponding algorithm exists in the analyzed program.

Other Methods. The main shortcoming of behavioral analysis approach for ransomware prevention is the potential false results due to the lack of an accurate decision mechanisms. In order to increase the accuracy of detection, anti-ransomware systems aim to consider more indicators which distinguish ransomware from benign applications. As the number of rules increases, simple decision techniques become inadequate. For this purpose, Machine Learning (ML) algorithms are used to analyze benign applications and known ransomware samples to extract feature vectors, build models and classify them. Recently, a ML based ransomware defense system has been made commercially available [21]. Meanwhile, the debate over the security of ML based malware defense systems continues. For instance, Hu and Tan proposed an algorithm to generate adversarial examples which cause the ML based malware detection systems to misclassify the applications [24].

Beside technical solutions, Lu and Liao suggest improving user awareness to help mitigate ransomware [32]. Security education for end users would effectively prevent ransomware attacks originating from phishing or spam emails. However, the attack surface that ransomware can exploit is far more larger. As the recent WannaCry attack demonstrates, ransomware evolution has enabled it to spread over the network. Especially, zero-day attacks can amplify the damage of ransomware and user education cannot help in this case.

3 Potential new threats

We start by giving high-level descriptions of advanced techniques that ransomware may utilize to defeat the defense systems characterized in the previous section. Next, we point out new areas that ransomware may exploit and extend the attack surface that next generation ransomware may target. In each discussion, our observations are supported by the real world incidents.

3.1 Rootkit-based Ransomware

Rootkit is a type of malware that has the ability to conceal its activities on the target computer system, e.g., code executions, file I/O, network and connections [22]. The capability of hiding malicious operations is achieved by hooking operating system's APIs in order to filter and remove the rootkit's traces, as depicted in Figure 1. Since a rootkit clears its footprints from APIs that inspect file and memory access, the rootkits are harder to detect than other types of malware.

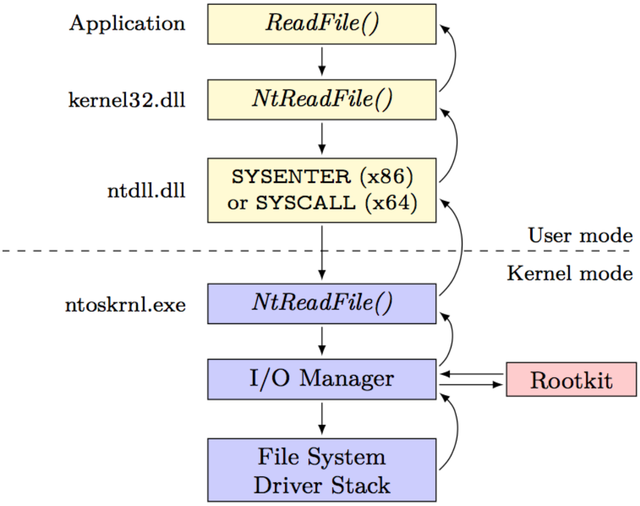


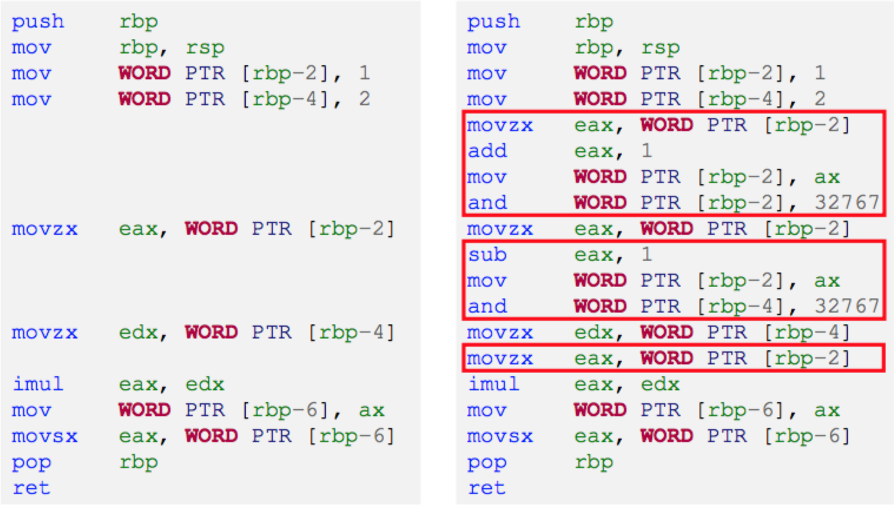
Figure 1: Interception of read calls by a kernel mode rootkit in order to hide its trace.

Hooking system APIs can be accomplished in several ways, including changing the function addresses in Import Address Table (IAT), patching System Service Dispatch Table (SSDT) in kernel level, and injecting code into applications (DLL injection) [39]. Starting from Windows Server 2003, x64-based versions of Windows platform introduced Kernel Patch Protection (KPP) which forces kernel mode drivers to be digitally signed, hence prevents unknown modification of code or critical structures in Windows kernel [34]. Nevertheless, cybercriminals frequently used stolen certificates to sign malware in order to penetrate this defense [9, 41]. Ransomware authors also seems to have this capability. A recent Virus- Total report shows that a sample of Razy ransomware has a valid digital signature [45].

Implementations of current ransomware defense approaches deeply rely on the security guarantees of the host OSes. While increasing the bar for cybercriminals, state-of-the-art ran- somware defense systems utilizes user mode hooks or kernel mode drivers to monitor behavior of applications and stop ransomware [13, 27, 28, 37]. Although there is currently no known ransomware which utilizes the advanced techniques of rootkits, the aforementioned defense systems may not detect a rootkit-based ransomware.

3.2 Obfuscation

Obfuscation is the practice of making a software implementation incomprehensible through a sequence of transformations while preserving the program semantics [12]. Originally, legitimate vendors utilized obfuscation to protect intellectual property in software implementation. However, malware authors also take advantage of obfuscation to conceal malicious executable code in the binary programs. Concordantly, obfuscated malware can evade from *signature based detection* techniques which is one of the oldest approaches in the battle with malware.



```
push    rbp
mov     rbp, rsp
mov     WORD PTR [rbp-2], 1
mov     WORD PTR [rbp-4], 2




movzx   eax, WORD PTR [rbp-2]



movzx   edx, WORD PTR [rbp-4]

imul    eax, edx
mov     WORD PTR [rbp-6], ax
movsx   eax, WORD PTR [rbp-6]
pop     rbp
ret
```

```
push    rbp
mov     rbp, rsp
mov     WORD PTR [rbp-2], 1
mov     WORD PTR [rbp-4], 2
movzx   eax, WORD PTR [rbp-2]
add     eax, 1
mov     WORD PTR [rbp-2], ax
and     WORD PTR [rbp-2], 32767
movzx   eax, WORD PTR [rbp-2]
sub     eax, 1
mov     WORD PTR [rbp-2], ax
and     WORD PTR [rbp-4], 32767
movzx   edx, WORD PTR [rbp-4]
movzx   eax, WORD PTR [rbp-2]
imul    eax, edx
mov     WORD PTR [rbp-6], ax
movsx   eax, WORD PTR [rbp-6]
pop     rbp
ret
```

Figure 2: Two code fragments that are semanti- cally equivalent and multiply the integers 1 and 2. Left, the original function. Right, the transformed function by adding ineffective instructions shown in- side red boxes. Note that the code's appearance is changed while keeping its behavior same.

Obfuscating malware can be categorized into four types: *encrypting*, *oligomorphic*, *polymorphic* and *metamorphic* mal- ware [48]. The members of the first type encrypts malicious code segment in the binary program and decrypt it in the runtime. This involves a decryptor function embedded in the malware body to decrypt and execute the malicious code. Anti-malware systems, though, would still recognize the decryptor function and identify malicious software. Thus, the second type, oligomorphic malware, carries a set of encrypted decryptors in data segment of binary and changes the decryptor in each generation. However, the number of decryptors is limited and therefore all of them eventually gets identified by anti-malware systems. On the other hand, polymorphic malware mutates its decryption engine randomly, hence evades signature based detection. The means of mutation include dead code insertion, register reassignment, subroutine reordering, instructor substitution, code transposition & integration. For instance, dead code insertion is the practice of adding code that has no effect on the functionality of the software and is shown in Figure 2. For the details of other techniques, we refer the reader to [2]. Anti-malware vendors developed *sandboxing* approach to help detection, which works by observing the program's behavior in a safe environment. Once the polymorphic malware is executed in sandbox and the constant malicious part is decrypted in the memory, signature based detection can be applied. The race between cybercriminals and anti-malware vendors resulted the appearance of metamorphic malware which actively recognizes, parses and mutates its whole body. As it does not contain a constant body, and thus cannot be detected via signature analysis [7], metamorphic malware has been considered to be most dangerous type.

In the ransomware side, the situation seems to be safe for now. As of today, there is no known instance of obfuscated ransomware through aforementioned techniques. Contemporary ransomware utilizes binary packers, *e.g.*, UPX[1], ASPack[2] or PEtite[3], which are used to compress the compiled code in order to make the size of executable even smaller. However, malware authors do not confine themselves to well-known packers, often write their own obfuscator routines and utilize combined packers [43]. This multi-layer protection may hinder defense systems based on API monitoring (if third party crypto libraries statically linked) and sandboxing. In the case of an unlucky event of infection, such a ransomware can be devastating.

3.3 White-Box Cryptography

White-box cryptography is the concept of protecting the sensitive data hard-coded in a software implementation [10, 11]. In particular, main focus of this domain is to embed secret keys into the source code in such a way that it is hard to extract them from compiled binary. An example of a Feistel network based block cipher and its fixed-key white-box implementation are illustrated in Figure 3. Although white-box cryptography is not a new idea (it is first introduced in 2002), no secure white-box implementation of the block cipher AES exists yet, for instance, previous proposals are found to be open to key extraction and table-decomposition attacks [5]. Nevertheless, white-box cryptography still continues to be an active field of research [3, 6, 25].

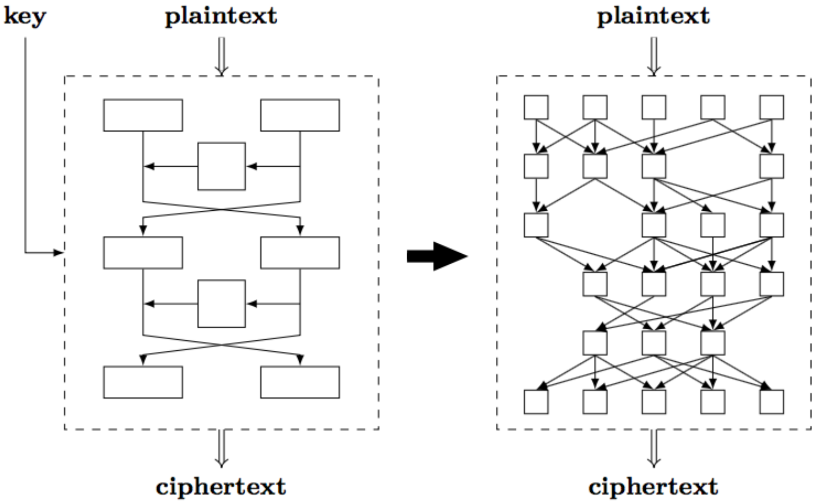


Figure 3: Left, a block cipher algorithm based on Feistel network structure. Right, a white-box implementation of that block cipher where a key is hard-coded into the algorithm.

Currently, ransomware implementations cannot protect the secret keys in the memory during the encryption process. Using this weakness, defense systems can extract these keys using various techniques. For instance, a key escrow like approach monitors calls to known cryptographic APIs (either built in or third party) and stores parameters of encryption functions in a vault [28]. In virtual environments, point-in-time snapshot of memory would also reveal those keys and recovery could be possible. Furthermore, some ransomware families encrypt victim's files using a key which is hard-coded in the ransomware body [29]. In this case, binary analysis can be utilized to search for static encryption keys in the compiled code. In other words, one can interact with the ransomware and propose solutions if the encryption keys resides unprotected in the memory. That being said, key extraction from securely implemented white-box algorithms is meant to be hard. Therefore, introducing of secure white- box implementations of block ciphers can tip the balance in favor of ransomware authors.

3.4 Ransomware of Things

Internet of Things (IoT) refers to the interconnected network of physical devices that can communicate over the Inter- net [20]. An IoT device can be equipped with electronic components, firmware, software, various types of sensors to collect information and actuators that allows to interact with the physical environment. Besides electronic devices like tele- visions, mobile phones and surveillance systems, in today's world, cars, planes, buildings, kitchen gadgets and even toys are also connected to the web.

IoT devices has been a part of our daily lives for a long time and can be seen virtually everywhere. However, IoT devices are inherently resource-constrained (CPU with low clock rate, small memory size). As such, the available options for cryptographic algorithms to use is limited when designing a secure communication protocol [49]. The security issues with IoT have always been a concern in information community [1], most importantly access control problems.

Given that the vulnerabilities in IoT devices and the high motivation of cyber criminals, there have already occurred several alarming and threatening ransomware incidents as follows. Hackers took control of ticket machines of San Fransisco's public transportation network and claimed ransom [47]. Furthermore in Austria, a hotel had to pay ransom after a ransomware infected its management system and blocked generating new cards [4]. Researchers demonstrated a proof- of-concept that the control of an Internet-enabled thermostat can be taken by a ransomware, allowing them to change the heating settings [42]. Similarly, A recent security report states that cybercriminals launched a Permanent Denial of Service (PDoS) attack on IoT devices which wipes all data on the device and destroy its firmware and/or basic functions, causing a permanent corruption [36].

By extending the attack surface and lack of adequate security, IoT has a potential of opening doors to novel ransomware attacks. For example, researchers demonstrated that it is possible to take control of a car and remotely stop it [18]. Also, another group of researchers showed that 75% of bluetooth smart door locks can be wirelessly hacked [46]. Given these facts, it is reasonable to ask the following questions: Consider that your car was remotely stopped in a rural area. *Would you pay the ransom to re-activate the car's engine?* Likewise, when you return your home in the middle of the night and see that your door is locked. *Would you pay the ransom to go in your home?* The picture may become worse for the enterprises, as the ransom amounts can be set higher and this makes the enterprises a more plausible target for cyber criminals. But the negative effects of a ransomware attack is beyond the money: the damage in the reputation and work loss should also be counted. Taking into the account that the security flaws in IoT devices do not seem to be fixed soon, or even fixable [38], ransomware attacks may gravitate towards IoT in the near future.

3.5 Socio Technical Attacks

The ultimate goal of cyber-criminals is to obtain money as much as possible. To achieve this, they can become very creative and employ novel marketing strategies. In one of these, a ransomware variant called Popcorn Time offers an option to victims who want to get decryption keys without paying. The condition is first victim infects other two ones and these two victims pay the ransom. Then, the first victim obtains the keys. The initial samples of Popcorn Time ransomware have an encryption key embedded in the malware body [14]. Although the key can be extracted from the current sample of Popcorn Time and files can be recovered for now, previous evolution of ransomware suggests that future samples of Popcorn Time may become more effective.

To this day, the vast majority of famous ransomware families share the same principle. Extortion by holding decryption keys can be expected to succeed when its vital for victims to regain access to their data. However, on the other side of medallion, there is another fact. Some data may need to be kept private such that when leaked, data owner may lose advantage and/or have economical damage. Thus, another way to extort victims can be to exfiltrate sensitive data and ask for a ransom to not make it public. These data types may include trading secrets, financial records, medical his- tory, government documents, details of high-tech projects, blue-prints of critical infrastructures, and internal/private communications. For example, the disclosure of data breaches reduced the purchase price of Yahoo by $350 million when it is acquired by Verizon [16]. It comes to mind that, instead of selling the leaked data in the underground market, hackers can try to claim a ransom to get a higher revenue. Another attack hit Sony Pictures, hackers compromised the computers and released sensitive data including company's financial records and e-mail messages of executives [23]. The contents of the breach put the company in a difficult situation so that one may ask the question: *Would Sony Pictures pay a ransom if attackers demand it?*

Lastly, we would like to point an important difference between extortion via encryption and data exfiltration. In the former case, the instance of threat comes to an end when the victims regain access to their files. In contrast, no one can guarantee that could retain cyber-criminals from asking for ransom again in the latter case. In this situation, it would be safe to expect that extortion via stealing sensitive information may be an increasing trend in the near future and prepare the network infrastructures against this threat.

4 Conclusion

Ransomware is a class of malware whose goal is to extort money, a goal that is facilitated by current anonymous currencies which guarantee to cyber-criminals to be paid without being traced. Then we need solid defense systems against what can easily degenerate in a pandemia of digital crimes. How- ever, unlike conventional anti-malware systems, ransomware mitigation does not tolerate mistake. If the ransomware is implemented properly and the attack succeeds, then the damage taken may be irreversible.

Existing ransomware mitigation systems are build upon the analysis of collected samples but a better strategy is to anticipate the future, and be prepared for the ransomware that will come. In this respect, we described possible threats that ransomware may pose by relying on novel techniques, like root-kit, obfuscation, and white-box, not yet adopted in real attack as well as by targeting critical domains, such as the Internet of Things and the Socio-Technical systems, which will worrisomely amplify the effectiveness of ransomware attacks. Our research is timely, since it is known that we must design products keeping security in mind, not integrating after whereas network infrastructures must be carefully configured and fully patched in order to prevent ransomware attacks through data exfiltration. We hope that our observations help developing and building more robust defense systems against ransomware threat.

Acknowledgements

This work is supported by a partner- ship between “pEp Security SA” and the Interdisciplinary Centre for Security, Reliability and Trust.

Notes

[1] Ultimate Packer for eXecutables, https://upx.github.io/
[2] ASPack, http://www.aspack.com/aspack.html
[3] PEtite, http://www.un4seen.com/petite/

References

[1] Luigi Atzori, Antonio Iera, and Giacomo Morabito. 2010. The Internet of Things: A survey. *Computer Networks* 54, 15 (2010), 2787 – 2805.

[2] Arini Balakrishnan and Chloe Schulze. 2005. Code obfuscation literature survey. (2005).

[3] Marc Beunardeau, Aisling Connolly, Remi Geraud, and David Naccache. 2016. White-box cryptography: Security in an insecure environment. *IEEE Security & Privacy* 14, 5 (2016), 88–92.

[4] Dan Bilefsky. 2017. Hackers Use New Tactic at Austrian Hotel: Locking the Doors. (30 Jan. 2017). Retrieved June 19, 2017 from https://www.nytimes.com/2017/01/30/world/europe/hotel-austria-bitcoin-ransom.html

[5] Andrey Bogdanov and Takanori Isobe. 2015. White-Box Cryptography Revisited: Space-Hard Ciphers. In Proc. 22nd ACM Conf. Comput. and Commun. Security (CCS ’15).

[6] Andrey Bogdanov, Takanori Isobe, and Elmar Tischhauser. 2016. Towards Practical Whitebox Cryptography: Optimizing Effi- ciency and Space Hardness. In *Proc. 22nd Int. Conf. Theory and Application of Cryptology and Inform. Security (ASIACRYPT 16)*.

[7] Jean-Marie Borello and Ludovic Mé. 2008. Code obfuscation techniques for metamorphic viruses. *Journal in Computer Virology* 4, 3 (2008), 211–220.

[8] Bromium. 2014. Understanding Crypto-Ransomware. (2014). Retrieved June 22, 2017 from https://www.bromium.com/sites/default/files/rpt- bromium-crypto-ransomware-us-en.pdf

[9] Thomas M. Chen and Saeed Abu-Nimeh. 2011. Lessons from stuxnet. *Computer* 44, 4 (2011), 91–93.

[10] Stanley Chow, Philip Eisen, Harold Johnson, and Paul C. Van Oorschot. 2003. White-Box Cryptography and an AES Implementation. In *Proc. Int. Workshop Select. Areas in Cryptography (SAC ’02)*.

[11] Stanley Chow, Phil Eisen, Harold Johnson, and Paul C. van Oorschot. 2003. A White-Box DES Implementation for DRM Ap- plications. In *Proc. ACM Workshop on Digital Rights Manage.* (DRM ’02).

[12] Christian Collberg, Clark Thomborson, and Douglas Low. 1998. Manufacturing Cheap, Resilient, and Stealthy Opaque Constructs. In *Proc. 25th ACM Symp. Principles of Programming Lan- guages (POPL ’98)*.

[13] Andrea Continella, Alessandro Guagnelli, Giovanni Zingaro, Giulio De Pasquale, Alessandro Barenghi, Stefano Zanero, and Federico Maggi. 2016. ShieldFS: A Self-healing, Ransomware- aware Filesystem. In *Proc. 32nd Annu. Conf. Comput. Security Applicat. (ACSAC ’16)*.

[14] Paul Ducklin. 2016. Popcorn Time ransomware lets you off if you infect two other people. (15 Dec. 2016). Retrieved July 13, 2017 from https://nakedsecurity.sophos.com/2016/12/15/popcorn-ti me-ransomware-lets-you-off-if-you-infect-two-other-people/

[15] Shona Ghosh. 2017. The massive global cyberattack affecting 200,000 victims will cause more chaos on Mon- day. (14 May 2017). Retrieved June 23, 2017 from http://uk.businessinsider.com/europol- said- there- are- 200000- cyberattack-victims-and-the-number-will-go-up-2017-5

[16] 2017. Verizon Will Pay $350 Million Less for Yahoo. (21 Feb. 2017). Retrieved July 13, 2017 from https://www.nytimes.com/2017/02/21/technology/verizon-will-pay-350-million-less-for-yahoo.html

[17] Dan Goodin. 2017. Web host agrees to pay $1m after it's hit by Linux-targeting ransomware. (6 June 2017). Retrieved June 20, 2017 from https://arstechnica.com/security/2017/06/web-host-agrees-to-pay-1m-after-its- hit-by-linux-targeting-ransomware/

[18] Andy Greenberg. 2015. Hackers Remotely Kill a Jeep on the Highway—With Me in It. (21 July 2015). Retrieved June 23, 2017 from https://www.wired.com/2015/07/hackers-remotely-kil l-jeep-highway/

[19] Felix Gröbert, Carsten Willems, and Thorsten Holz. 2011. Automated Identification of Cryptographic Primitives in Binary Programs. In *Proc. 14th Int. Conf. Recent Advances in Intrusion Detection (RAID '11)*.

[20] Jayavardhana Gubbi, Rajkumar Buyya, Slaven Marusic, and Marimuthu Palaniswami. 2013. Internet of Things (IoT): A vision, architectural elements, and future directions. *Future generation computer systems* 29, 7 (2013), 1645–1660.

[21] Peter Hale. 2017. Acronis True Image 2018: Artificial Intelligence Meets Intelligent Backup. (30 Aug. 2017). Retrieved October 2, 2017 from https://www.acronis.com/en-us/blog/posts/acronis-true-image-2018- artificial- intelligence-meets-intelligent-backup

[22] Greg Hoglund and James Butler. 2006. *Rootkits: subverting the Windows kernel*. Addison-Wesley Professional.

[23] Amanda Holpuch. 2014. Sony email hack: what we've learned about greed, racism and sexism. (15 Dec. 2014). Retrieved July 13, 2017 from https://www.theguardian.com/technology/2014/ dec/14/sony-pictures-email- hack-greed-racism-sexism

[24] Weiwei Hu and Ying Tan. 2017. Generating Adversarial Malware Examples for Black-Box Attacks Based on GAN. https://arxiv.org/abs/1702.05983. (2017).

[25] Yin Jia, TingTing Lin, and Xuejia Lai. 2016. A generic attack against white box implementation of block ciphers. In *Proc. Int. Conf. Comput. Inform. and Telecommun. Systems (CITS '16)*.

[26] Kaspersky. 2016. Security Bulletin 2016. (Dec. 2016). Re- trieved June 22, 2017 from https://securelist.com/files/2016/12/KSB2016_Story_of_the_Year_ENG.pdf

[27] Amin Kharaz, Sajjad Arshad, Collin Mulliner, William Robertson, and Engin Kirda. 2016. UNVEIL: A Large-Scale, Automated Approach to Detecting Ransomware. In *Proc. 25th USENIX Security Symp. (USENIX Security '16)*.

[28] Eugene Kolodenker, William Koch, Gianluca Stringhini, and Manuel Egele. 2017. PayBreak: Defense Against Cryptographic Ransomware. In *Proc. ACM Asia Conf. Comput. and Commun. Security (ASIACCS '17)*.

[29] Ondrej Kubovič. 2016. Ransomware is everywhere, but even black hats make mistakes. (28 April 2016). Retrieved June 19, 2017 from https://www.welivesecurity.com/2016/04/28/ransom ware- is- everywhere- but- even-black- hats- make- mistakes/

[30] Security Ledger. 2015. FBI's Advice on Ransomware? Just Pay The Ransom. (22 Oct. 2015). Retrieved June 22, 2017 from https://securityledger.com/2015/10/fbis- advice- on-cryptol ocker-just-pay-the-ransom/

[31] Pierre Lestringant, Frédéric Guihéry, and Pierre-Alain Fouque. 2015. Automated Identification of Cryptographic Primitives in Binary Code with Data Flow Graph Isomorphism. In *Proc. 10th ACM Symp. Information Comput. and Commun. Security (ASIACCS '15)*.

[32] Xin Luo and Qinyu Liao. 2007. Awareness Education as the Key to Ransomware Prevention. *Information Systems Security* 16, 4 (2007), 195–202.

[33] Trend Micro. 2016. Kansas Hospital Hit by Ransomware, Extorted Twice. (23 May 2016). Retrieved June 23, 2017 from https://www.trendmicro.com/vinfo/us/security/news/cybercri me- and- digital-threats/kansas-hospital-hit-by-ransomware-extorted-twice

[34] Microsoft. 2007. Kernel patch protection: frequently asked questions. (Jan. 2007). Retrieved June 13, 2017 from https://msdn.microsoft.com/en- us/library/windows/hardware/Dn613955(v=vs.85).aspx

[35] Satoshi Nakamoto. 2008. Bitcoin: A peer-to-peer electronic cash system. https://bitcoin.org/bitcoin.pdf. (2008).

[36] Radwire. 2017. "BrickerBot" Results In PDoS At- tack. (4 May 2017). Retrieved June 23, 2017 from https://security.radware.com/ddos- threats- attacks/bricke rbot-pdos-permanent-denial-of-service/

[37] Nolen Scaife, Henry Carter, Patrick Traynor, and Kevin R.B. Butler. 2016. CryptoLock (and Drop It): Stopping Ransomware Attacks on User Data. In *Proc. 36th Int. Conf. Distributed Computing Syst. (ICDCS '16)*.

[38] Bruce Schneier. 2014. The Internet of Things Is Wildly Insecure – And Often Unpatchable. (6 Jan. 2014). Retrieved June 23, 2017 from https://www.wired.com/2014/01/theres-no-good-way-to- patch- the- internet- of- things- and- thats- a- huge- problem/

[39] Spencer Smith and John Harrison. 2012. Rootkits. (2012). Retrieved June 13, 2017 from http://www.symantec.com/content /en/us/enterprise/media/security_response/whitepapers/rootki ts.pdf

[40] Symantec. 2016. An ISTR Special Report: Ransomware and Businesses 2016. (19 July 2016). Retrieved June 20, 2017 from https://www.symantec.com/content/en/us/enterprise/media/security_response/whitepapers/ISTR2016_Ransomware_and _Businesses.pdf

[41] Peter Szor. 2011. Duqu–Threat Research and Analysis. (Nov. 2011). Retrieved June 13, 2017 from https://securingtomorrow.mcafee.com/wp- content/uploads/2011/10/Duqu.pdf

[42] Andrew Tierney. 2016. Thermostat Ransomware: a lesson in IoT security. (Aug. 2016). Retrieved June 19, 2017 from https://www.pentestpartners.com/security-blog/thermostat-ransomware-a-lesson-in-iot-security/

[43] Xabier Ugarte-Pedrero, Davide Balzarotti, Igor Santos, and Pablo G. Bringas. 2015. SoK: deep packer inspection: a lon- gitudinal study of the complexity of run-time packers. In *Proc. 36th IEEE Symp. on Security and Privacy (S&P '15)*.

[44] Cybersecurity Ventures. 2017. Ransomware Damage Report. (18 May 2017). Retrieved June 23, 2017 from http://cybersecurityv entures.com/ransomware-damage-report-2017-5-billion/

[45] VirusTotal. 2017. Scan report. (15 June 2017). Retrieved July 13, 2017 from https://virustotal.com/en/file/81fdbf04f3d0d9a85e0f bb092e257a2dda14c5d783f1c8bf3bc41038e0a78688/analysis/

[46] Paul Wagenseil. 2016. 75 Percent of Bluetooth Smart Locks Can Be Hacked. (Aug. 2016). Retrieved June 19, 2017 from http://www.tomsguide.com/us/bluetooth-lock-hacks-defc

[47] Elizabeth Weise. 2016. Ransomware attack hit San Francisco train system. (28 Nov. 2016). Retrieved June 22, 2017 from https://www.usatoday.com/story/tech/news/2016/11/28/san-francisco-metro-hack-meant-free-rides-saturday/94545998/on2016,news-23129.html

[48] lsun You and Kangbin Yim. 2010. Malware Obfuscation Tech- niques: A Brief Survey. In *Proc. 5th Int. Conf. Broadband,* Wireless Computing, Commun. and Applicat. (BWCCA '10).

[49] Kai Zhao and Lina Ge. 2013. A Survey on the Internet of Things Security. In *Proc. 9th Int. Conf. Computational Intelligence* and Security (CIS '13).

ADVANCES IN CYBERSECURITY 2017
I. Bernik, B. Markelj & S. Vrhovec

Children Privacy Protection in Video Surveillance Based on Automatic Age Estimation

PETRA GRD, IGOR TOMIČIĆ & MIROSLAV BAČA

Abstract Video surveillance can be defined as using video cameras to observe an area. These systems can have many benefits, however there are also many risks. The problem mentioned most often is the violation of privacy. One of the most vulnerable groups whose privacy needs to be protected are children. This paper analyses the importance of children privacy protection in video surveillance and proposes a model for children privacy protection based on age estimation. The proposed model uses automatic age estimation in order to distinguish between children and adults. Age can be estimated in different ways, the model proposed in this paper classifies people by using their face anthropometry.

Keywords: • privacy • children • video surveillance • face • body • age estimation • anthropometry •

CORRESPONDENCE ADDRESS: Petra Grd, Ph.D., Assistant Professor, University of Zagreb, Faculty of Organization and Informatics, Pavlinska 2, 42000 Varaždin, Croatia, e-mail: petra.grd@foi.hr. Igor Tomičić, Ph.D., Senior Assistant, University of Zagreb, Faculty of Organization and Informatics, Pavlinska 2, 42000 Varaždin, Croatia, e-mail: tomicic.igor@gmail.com. Miroslav Bača, Ph.D., Full Professor, University of Zagreb, Faculty of Organization and Informatics, Pavlinska 2, 42000 Varaždin, Croatia, e-mail: mbaca@foi.hr.

DOI https://doi.org/10.18690/978-961-286-114-8.9 ISBN 978-961-286-114-8

Available at: http://press.um.si.

1 Introduction

People go through different changes during their growth and aging. Most of the changes during a persons aging process are changes in body appearance, especially craniofacial morphology. Different craniofacial characteristics appear at different age and change during the aging process. Based on these changes, age of a person can be automatically estimated. Automatic age estimation has numerous applications, many of them in security. This paper proposes a model for another application of age estimation, for children privacy protection in today's world.

In the recent years there have been many terrorist threats world-wide and an increase in criminal rates, especially in urban areas. This resulted with the increase of video surveillance systems in both public places such as airports, streets, banks and in private houses [5]. Video surveillance is defined as "a system of monitoring activity in an area or building using a television system in which signals are transmitted from a television camera to the receivers by cables or telephone links forming a closed circuit [25]. Video surveillance has many benefits, but it also has many challenges. One of the challenges which is widely discussed is the importance and lack of privacy in video surveillance often referred to as privacy violation. There are many different definitions and aspects of privacy. In connection to public video surveillance, privacy can be defined as the "ability to prevent other parties from learning one's current identity by recognising his/her personal characteristics" [36]. Privacy is one of the fundamental human rights and needs to be protected. The question that arises is what and who should be protected. The aim of human rights instruments [28] is the protection of those vulnerable to violations of their fundamental human rights. There are particular groups who are vulnerable or have traditionally been victims of violations and require special protection for the equal and effective enjoyment of their human rights [19]. Vulnerability is the degree to which a population, individual or organization is unable to anticipate, cope with, resist and recover from the impacts of disasters [29]. This paper focuses on children as one of the most vulnerable groups whose privacy needs to be protected. Children (minors) are defined as people from birth to age seventeen.

The idea of this paper is to propose a model for automatic children privacy protection in video surveillance. Most of the existing research focuses on different types of hiding privacy information in videos. This paper focuses on the issue of video analysis and finding regions of interest, which is the precursor to hiding privacy information. To this end, the model proposes usage of face detection and automatic age estimation based on face images.

The rest of the paper is structured as follows; in section 2, we present an overview of related research, focusing primarily on public video surveillance aspects and on automatic age estimation methods. Section 3 presents proposed model, and section 4 provides conclusions, final thoughts, and plans on future research on the subject.

2 Related work

The related literature overview within this section presents the most relevant aspects in existing body of research dealing with the nature, technology and policies of public video surveillance, the questions of children privacy in public video surveillance systems, and an overview of research efforts dealing with automatic age estimation.

2.1 Public Video Surveillance Overview

Most of the active surveillance systems are non-discriminative by nature of their implementation and usage, which means that they are surveying everyone, possibly posing a threat to human privacy, rights and individual freedoms [2, 34].

As Korshunov and Ebrahimi argue, the latest progress in video analytics such as detection, recognition and tracking, combined with personal data available from web and social networks, are indicative parts of the emerging multi-modal surveillance systems, which "pose a serious threat to fundamental rights to privacy." [21]

Various approaches were taken in an effort to preserve the privacy of individuals under video surveillance. For example, in [32] people's identities are protected by irreversibly masking their faces with a colored ellipse. A step further is argued in [1], where authors propose a method for obscuring the whole body silhouette. Korshunov and Ebrahimi [21] give detailed overview of the existing methods and techniques for privacy protection, but also argue about their shortcomings; authors propose that the following properties should be incorporated into a practical privacy protection method: low complexity, reversibility, flexibility of application, security and variable strength granularity. With these properties in mind, authors further propose a geometrical transformation (warping) for protection of visual privacy.

The Urbaneye Project [17] has been documenting the proliferation of surveillance cameras in public and semi-public spaces in Europe and, at the time of research, showed that 29 percent of publicly accessible institutions used a certain form of video surveillance. Such data may point to the necessity of having laws that apply explicitly to video surveillance systems.

2.2 Legal Aspects on Public Video Surveillance in Different Countries

Rajpoot and Jensen [31] reported on the legislation statuses regarding privacy in video surveillance systems in Canada, United States, and other selected countries in Europe. For example, in most of the analysed countries, rights to privacy are not explicitly mentioned in constitutions; only Denmark, Netherlands and Spain have the right to privacy formally recognised within constitutions. In most countries however, it is obligatory to inform public about video surveillance via signs, and most of the countries do have some form of regulatory bodies (in France, Netherlands and Spain for example, it is obligatory to report video surveillance to these bodies; some, like Norway, make a distinction between recorded and non-recorded surveillance), and defined data retention periods. Those periods range from vaguely descriptive ("as long as necessary and must be destroyed when no more required", as an example from Canada), derived from other laws (retention period for "other personal information" for 12 months is defined in Denmark and France), explicitly stated (up to four weeks in Netherlands, one month in Spain), to non-existing retention laws (United States, Norway).

Within the analysed literature, the questions of children privacy in public video surveillance are not explicitly dealt with.

2.3 Authomatic Age Estimation

Age estimation can be defined as the determination of the age of the person or his/her age group [11, 33]. Age can be estimated in different ways, however most of the papers focus on age estimation of humans from face images.

The algorithms most often used for age estimation are face age estimation algorithms and they have two basic parts: face representation and aging function learning method [12].

2.3.1 Representation

There are many different face representation models but five main types are recognised in literature [8, 9, 12, 15]: anthropometric model, active appearance model, aging pattern subspace, age manifold and biologically-inspired models.

Kwon and Lobo [22] in 1999 proposed the first algorithm for automatic age estimation from face images. Computations in this model are based on the craniofacial development theory. Changes in the appearance of face caused by growth are sufficient to categorize faces in several age groups. Their anthropometric model uses six ratios to distinguish between different age groups.

Active appearance model [4] was first expanded to age estimation by [23] suggesting an aging function defined by $age = f(b)$, to explain the variation in years. Age is the age of a person in the picture, b is a vector containing 50 parameters learned from AAM, and f is an aging function.

An automatic age estimation method named AGES (AGing pattErn Subspace) is proposed in [10]. Authors define aging pattern "as the sequence of a particular individual’s face images sorted in time order, by constructing a representative subspace", and model it.

Age manifold [8] is more flexible than AGES model and it allows to learn the common pattern of aging for more than one person at different ages instead of learning the specific aging pattern for each person.

One of the most accurate age estimation algorithms to date is the EBIF [6] algorithm based on biologically inspired models. The idea for biologically inspired model (BIF) came from human vision system. It showed good results in object recognition, so Guo et al. [15] adapted this model to human age estimation based on face images.

2.3.2 Aging function learning

Aging function learning methods can be viewed as [15] a multiclass classification problem and a regression problem.

In age estimation as classification problem, each age label is treated as a single class [8]. Multiclass classification can further be divided [3] into age-group classification, single-level age estimation and hierarchical age estimation. Age-group classification roughly estimates the age group a person belongs to. In single-level age estimation each age is viewed as one class. Hierarchical age estimation first finds the age group a person belongs to, and second step is to find the exact age in this age group [3]. There are a large number of classifiers used in age estimation [3, 14]: Artificial Neural Networks, Support Vector Machines, Nearest Neighbour, Quadratic function, Fuzzy Linear Discriminant Analysis, Hierarchical estimation, Multilayer Perceptron...

If age is observed as a sequential set of values, age estimation can be viewed as a regression problem [8]. Regressors used for age estimation can be [14]: Quadratic Function, Multiple Linear Regressor, Support Vector Regression, Semi-definite Programming Technique, Expectation-Maximization, Robust Multi-instance Regression, Least Angle Regression...

Current research results for age classification can be seen in Table 1. Further research and development will improve the results. Also, there are not many papers that research classification into minors and adults.

3 The proposed model

In Figure 1, the most common video surveillance system architecture can be seen. The model proposed in this paper should be deployed on the surveillance server. On the same server, the original video recordings are stored. After the privacy protection, the privacy protected videos are the only ones authorised persons with lower security clearance have access to. Access to the original video recordings should be possible with special permission or to someone with high security clearance.

Table 1: Research results for age classification

Paper	Classes	Accuracy
[18]	0-2, 3-39, 40-59, 60+	81.58%
[26]	0-12, 13-18, 19-59, 60+	94.28%
[37]	Youths and adults	86%
[30]	Kids and adults	90.46%
[16]	0-15, 16-30, 31-50, 50+	87.03%
[24]	0-2, 3-7, 8-12, 13-19, 20-36, 37-65, 66+	48.5%
[20]	0-2, 3-12, 13-24, 25-40, 41-55, 56+	95%
[35]	Baby and adult	49.72%
[13]	10+-5, 20+-5, 30+-5, 40+-5, 50+-5, 60+-5	80%

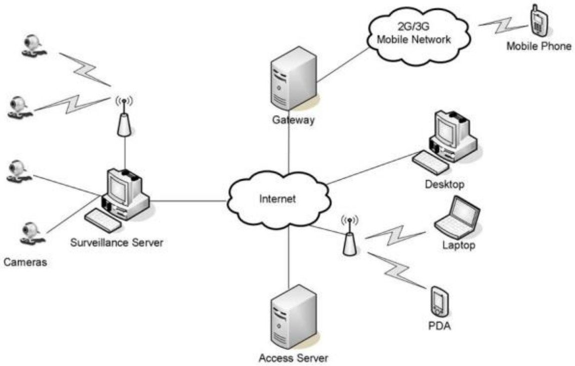


Figure 2: Video surveillance system architecture [5]

The idea of the proposed model is to detect humans in video surveillance and classify them into two groups: minors and adults. After the classification, images of minors in video recordings are scrambled in order to protect their privacy.

3.1 Model Description

The model consists of two main parts: analysis and information hiding. The part this model focuses on is video analysis (Figure 4). The age estimation part of the model was developed by [12]. First step is to obtain the image from surveillance camera, and preprocess the image. Image preprocessing includes noise reduction, image sharpening (if necessary), and detection of face in an image. The output from this part is the region of interest (ROI) in the video (face).

The ROI from the first step is the input in the feature extraction step. Different face representation models have been described in section 2. This model uses a modification of anthropometric model for face representation. In this step characteristic points on human face are detected. The positions of these points are the output from the feature extraction step. Different authors define a different number of characteristic points for age estimation. This model uses 26 face landmarks defined by [12] (Figure 2).

After the feature extraction, the ratios important for age classification need to be determined and calculated. The modification of the anthropometric model is visible in this step.

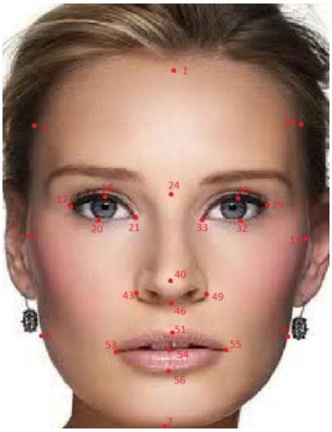


Figure 3: Landmark points used in this research [12]

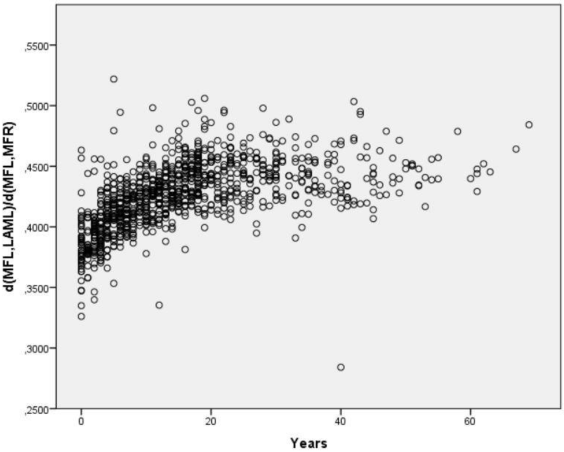


Figure 4: Correlation between years and one ratio [12]

Anthropometric model defines six ratios, and Grd [12] defines 62 ratios. Ratios are calculated as ratios of Euclidean distances (1) between face features.

$$d(A,B) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \tag{1}$$

All possible ratios are calculated (52650), and important ratios are defined based on statistical analysis of correlation between the calculated ratios on human face and age of a person. Scatter matrices are plotted which shows that correlation between ratios and age is non-linear (Figure 3).

Next step is to calculate the correlation between all the ratios and age, to this end Spearman coefficient is used. After calculation of Spearman coefficient, only 62 ratios with high correlation are selected for further calculation. Other ratios have medium or low correlation. The output of this step are the ratios values.

After ratios calculation, the next step is classification of videos into those of minors and adults. Classification in the algorithm is done with neural networks. More precisely, multi-layer perceptron is used. Input into this step are the ratios from the previous step and trained classifier. Dependent variable which needs to be determined is variable Class.

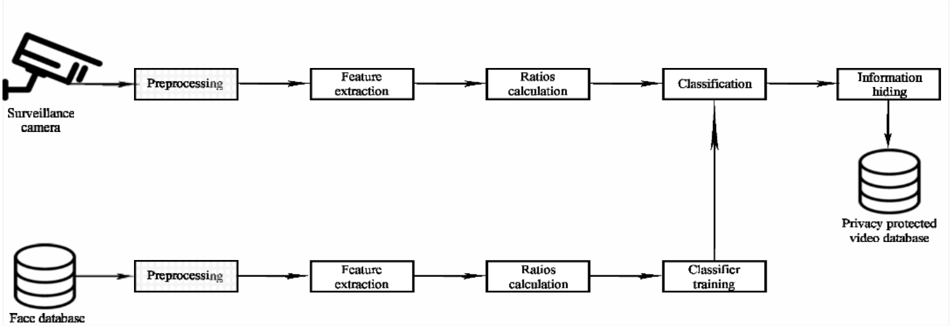


Figure 5: Model block diagram

There are two possible values: minors and adults. This variable is predicted using 62 covariates (ratios from previous step). Covariates are rescaled using Standardized method. The distribution mean and standard deviation for each feature need to be calculated. Than the mean is substracted from each covariate, and values of each feature are divided by its standard deviation (2), where x is a starting covariate value, x' is a new covariate value, mean is a distribution mean value, and s is the standard deviation of distribution [12].

$$x' = \frac{x - mean}{s} \tag{2}$$

After the decision is made, if the person in the video is a minor, the output of this step is video of a minor which information needs to be hidden. The final step is information hiding, and storing the privacy protected video in a database.

3.2 Model Performance

In order to assess the performance of the classifier, different measures are calculated. Measures most often used are accuracy (3), precision (4), recall (5) and specificity (6). Accuracy is the proportion of the total number of predictions that were correct, Precision is the proportion of positive cases that were correctly identified, Recall is the proportion of actual positive cases which are correctly identified and Specificity is the proportion of negative cases that were correctly identified [27].

$$Accuracy = \frac{TP+TN}{N} * 100\% \tag{3}$$

$$Precision = \frac{TP}{TP+FP} * 100\% \tag{4}$$

$$Recall = \frac{TP}{TP+FN} * 100\% \tag{5}$$

$$Specificity = \frac{TN}{FP+TN} * 100\% \tag{6}$$

TP is the number of correct predictions that an instance is positive, *TN* is the number of correct predictions that an instance is negative, and *N* is the total number of cases.

The classifier was tested on the 1002 images from FG NET data- base [7] and the calculated overall Accuracy was 81.34%, Precision was 86.47%, Recall was 83.91% and Specificity was 76.80%.

4 Conclusion and further research

The goal of this paper was to propose a model for privacy protection of minors in public surveillance videos, which are mostly non-discriminative by nature, and, as shown in subsection 2.2, rights to privacy are not explicitly mentioned in constitutions of most researched countries. The first part of the paper emphasised the importance of privacy in today's world where video surveillance is the norm, and not the exception. The emphasis was on the privacy protection of children as one of the most vulnerable groups. In section 2, the state of the art research in public surveillance and age estimation has been presented.

Most of the existing body of research on privacy protection in video surveillance context focuses on different ways of information hiding. The precursor step is the answer to the question - what, or who, needs to be hidden. This paper detects people in video surveillance, and uses automatic age estimation to classify images into those of minors or adults to answer that question. In order to directly address the issue of children's privacy, the next step in the proposed model describes the method of privacy protection - scrambling the images of minors in video recordings, while keeping the original recordings on the same surveillance server. While people with regular clearances would have access to the scrambled videos, only persons with special clearances would have access to the original recordings. Proposed model is described in detail in section 3.1.

Future research will focus on using multimodal biometrics, more precisely, face images and body images, and fusing the results in order to achieve better performance of the classifier. Also, future work will in more detail research surveillance server security and information hiding.

References

[1] Datong Chen, Yi Chang, Rong Yan, and Jie Yang. 2009. Protecting personal identification in video. *Protecting Privacy in Video Surveillance* (2009), 115–128.

[2] Simon Chesterman. 2010. Privacy and surveillance in the age of terror. *Survival* 52, 5 (2010), 31–46.

[3] Lee Y.J. Lee S.J. Park K.R. Kim J. Choi, S.E. 2011. Age Estimation Using a Hierarchical Classifier Based on Global and Local Facial Features. *Pattern recognition* (2011).

[4] Edwards G.J. Taylor C.J. Cootes, T.F. 1998. Active Appearance Models. *Proceedings of European Conference on Computer Vision* (1998).

[5] Frederic Dufaux, Mourad Ouaret, Yousri Abdeljaoued, Alfonso Navarro, Fabrice Vergnenegre, and Touradj Ebrahimi. 2006. Privacy enabling technology for video surveillance. *Proceedings of SPIE* 6250 (2006), 62500M–62500M–12.

[6] Al-Saban M. El Dib, M.Y. 2010. Human Age Estimation Using Enhanced Bio-inspired Features (EBIF). *International Conference on Image Processing* (2010).

[7] Fg-net. 2014. The Fg-net Aging Database. http://www-prima.inrialpes.fr/FGnet/html/benchmarks.html. (July 2014).

[8] Guo G.-Huang T.S. Fu, Y. 2010. Age Synthesis and Estimation via Faces: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2010), 1955–1976.

[9] Fu Y.-Smith-Miles K Geng, X. 2010. Automatic Facial Age Estimation. *Pacific Rim International Conferences on Artificial Intelligence* (2010).

[10] Xin Geng, Zhi-Hua Zhou, and Kate Smith-Miles. 2007. Automatic age estimation based on facial aging patterns. *IEEE Transactions on pattern analysis and machine intelligence* 29, 12 (2007), 2234–2240.

[11] Petra Grd. 2013. Introduction to Human Age Estimation Using Face Images. *Research papers Faculty of Materials Science and Technology Slovak University of Technology in Trnava* (2013), 35–41.

[12] Petra Grd. 2015. Two-dimensional face image classification for distinguishing children from adults based on anthropometry. *PhD thesis at the Faculty of organization and informatics, University of Zagreb, Croatia* (2015).

[13] A. Gunay and V.V. Nabiyev. 2013. Age Estimation Based on Local Radon Features of Facial Images. *Computer and Information Sciences III* (2013).

[14] G. Guo. 1994. Human Age Estimation and Sex Classification. *Video Analytics for Business Intelligence* (1994).

[15] Mu G.-Fu-Y. Huang T.S. Guo, G. 2009. Human Age Estimation Using Bio-inspired Features. *Conference on Computer Vision and Pattern Recognition* (2009), 112–119.

[16] Ebrahimnezhad H. Hajizedah, M.A. 2011. Classification of Age Groups from Facial Image Using Histograms of Oriented Gradients. *7th Iranian Machine Vision and Image Processing* (2011).

[17] Leon Hempel and E Töpfer. 2002. Urban Eye: Inception Report to the European Commission, 5th Framework Programme. *Technical University Berlin* (2002).

[18] Lee C.P.-Chen-C.W. Horng, W.B. 2001. Classification of Age Groups Based on Facial Features. *Tamkang Journal of Science and Engineering* (2001).

[19] Icelandic human rights center. 2017. The human rights protection of vulnerable groups. http://www.humanrights.is/en/human-rights-education-project/human-rights-concepts-ideas-and-fora/the-human-rights-protection-of-vulnerable-groups/. (September 2017).

[20] D. Kalamani and P. Balasubramanie. 2006. Age Classification using Fuzzy Lattice Neural Network. *Proceedings of the Sixth International Conference on Intelligent Systems Design and Application* (2006).

[21] Pavel Korshunov and Touradj Ebrahimi. 2013. Using warping for privacy protection in video surveillance. In *Digital Signal Processing (DSP), 2013 18th International Conference on*. IEEE, 1–6.

[22] Y.H. Kwon and N.V. Lobo. 1999. Age Classification from Facial Images. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (1999).

[23] Taylor C.-Cootes-T. Lanitis, A. 1998. Toward Automatic Simulation of Aging Effects on Face Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (1998).

[24] Liu Q.-Liu-J. Lu H. Li, C. 2012. Learning Distance Metric Regression for Facial Age Estimation. *21st International Conference on Pattern Recognition* (2012).

[25] Harper Collins Publishers Limited. 2017. Video Surveillance definition. webpage. (July 2017).

[26] J. Nityashri and G. Kulanthaivel. 2012. Classification of Human Age based on Neural Network Using Fg-net Aging Database and Wavelet. *IEEE Fourth International Conference on Advanced Computing* (2012).

[27] Lecture Notes. 2015. Precision and Recall. http://www.cs.odu.edu/mukka/cs495s13/Lecturenotes/Chapter5/recallprecision.pdf. (February 2015).

[28] OHCHR. 2017. International Human Rights Law. http://www.ohchr.org/EN/ProfessionalInterest/Pages/InternationalLaw.aspx. (September 2017).

[29] World Helath Organization. 2017. Vulnerable groups. http://www.who.int/environmental_health_emergencies/vulnerable_groups/en/. (September 2017).

[30] H. Qi and Zhang L. 2009. Age Classification System with ICA Based Local Facial Features. *Advances in Neural Networks, Lecture Notes in Computer Science* (2009).

[31] Qasim Mahmood Rajpoot and Christian Damsgaard Jensen. 2015. Video surveillance: Privacy issues and legal compliance. *Promoting Social Change and Democracy Through Information Technology* (2015), 69.

[32] Jeremy Schiff, Marci Meingast, Deirdre K Mulligan, Shankar Sastry, and Ken Goldberg. 2007. Respectful cameras: Detecting visual markers in real-time to address privacy concerns. In *Intelligent Robots and Systems, 2007. IROS 2007. IEEE/RSJ International Conference on*. IEEE, 971–978.

[33] Scholarpedia. 2017. Facial Age Estimation. webpage. (July 2017).

[34] Barrie Sheldon. 2011. Camera surveillance within the UK: Enhancing public safety or a social threat? *International Review of Law, Computers & Technology* 25, 3 (2011), 193–203.

[35] L.L. Shen and Z. Ji. 2008. Modelling Geometric Features for Face Based Age Classification. *roceedings of the Seventh International Conference on Machine Learning and Cybernetics* (2008).

[36] Takashi Koshimizu Naoko Nitta Yoshimichi Ito Xiaoyi Yu, Kenta Chinomi and Noboru Babaguchi. 2008. Privacy protecting visual processing for secure video surveillance. *Image Processing, 2008. ICIP 2008 International Conference on* (2008), 1672–1675.

[37] Miller P. Zhou, H. and J. Zhang. 2011. Age classification using Radon transform and entropy based scaling SVM. *Proceedings of the British Machine Vision Conference* (2011).

Security of IoT Cloud Services - A User-Oriented Test Approach

MARTIN BÖHM, INA SCHIERING & DIEDERICH WERMSER

Abstract The trend of digitalization fostered by Internet of Things technologies is characterized by a huge potential for innovations on one side and security and privacy threats on the other side. The emerging IoT cloud services offer great opportunities, since the complexity of providing IoT services is significantly reduced. On the other hand the level of security and privacy is not transparent to the users of the service beside the provided documentation. Based on a generalized architecture of IoT cloud services, the possibilities of testing security and privacy requirements of IoT cloud services from a user perspective are investigated. The proposed test framework is used to evaluate the IoT cloud services of the providers Amazon, Microsoft, Google and ThingSpeak.

Keywords: • IoT • IoT cloud service • cloud security • security testing • cloud privacy •

CORRESPONDENCE ADDRESS: Martin Böhm, M.S., Ostfalia University of Applied Sciences, Research Group IP-Based Communication Systems, Salzdahlumer Straße 46/48, 38302 Wolfenbüttel, Germany, e-mail: ma.boehm@ostfalia.de. Ina Schiering, Ph.D., Professor, Ostfalia University of Applied Sciences, Institute of Information Engineering, Straße 46/48, 38302 Wolfenbüttel, Germany, e-mail: i.schiering@ostfalia.de. Diederich Wermser, Ph.D., Professor, Ostfalia University of Applied Sciences, Research Group IP-Based Communication Systems, Salzdahlumer Straße 46/48, 38302 Wolfenbüttel, Germany, e-mail: d.wermser@ostfalia.de.

DOI https://doi.org/10.18690/978-961-286-114-8.10 ISBN 978-961-286-114-8

Available at: http://press.um.si.

1 Introduction

The Internet of Things (IoT) is an important paradigm connected to the trend of digitalization. IoT is based on interconnected devices, sensors resp. actors with limited storage and processing capacity, communicating via networks with a back-end service. This enables ubiquitous computing scenarios. Examples of such connected devices are temperature or light sensors and IP cameras. Gubbi et al. [19] state that the main aim of IoT is to “make a computer sense information without the aid of human intervention”. This emerging paradigm leads to innovations in areas as energy grids, city infrastructures, home automation and production systems.

Because of the restricted resources of IoT devices, data storage and computation is typically performed in a back-end system. An emerging paradigm in this area are IoT cloud services (see Botta et al. for an overview [9]). Traditional cloud providers as Google, Amazon, Microsoft, IBM, Oracle, Cisco etc. offer specialized services to connect smart devices. In addition there are projects resp. specialized providers as e.g. FIWARE, ThingSpeak, Heroku, OpenIoT, SensorCloud, Nimbits, Xively, Particle, CloudPlugs, Stack4Things.

An important issue in IoT are security and privacy challenges. Especially vulnerable smart home devices as for example IP cameras connected to the Internet, got infected due to weak or hard-coded passwords. Because of the homogeneity of devices such vulnerabilities can lead to well scaling attacks. The botnet Mirai consisting of such vulnerable devices attacked the Domain Name Service (DNS) provider Dyn “involving 10s of millions of IP addresses” [39]. Companies as Twitter, Amazon, Tumblr, Reddit, Spotify and Netflix were only partly reachable for the day of the attack [26]. An overview of security and privacy challenges in IoT is given by Sadeghi et al. [34] and Zhao et al. [41].

Beside vulnerabilities of IoT devices, according to Zhao et al. [41], also the network and application layers constitute security challenges. An example for vulnerabilities of the network layer in a smart city environment was described by Cerrudo [10] where a traffic control system did not use standard security mechanisms as encryption or authentication for the network connection to the back-end. Eyal Ronen et al. [32], [33] implemented an IoT worm for the popular Philips Hue smart bulbs. Because of an attack with ransomware e.g. the San Francisco Muni system was not able to sell tickets to passengers [16].

IoT Cloud Services constitute the opportunity to reduce the complexity of providing IoT services for users lacking the required expertise as IT service providers. But as a consequence these users rely on the security and privacy measurements of the IoT cloud service provider which are not transparent to users of the services. A further general risk for standard cloud services is the risk of vendor lock-in which is also present for IoT cloud services, because of the proprietary incompatible platforms. A *user* in this scenario is the organization which uses the cloud service to realize an IoT scenario.

Therefore we propose in this paper approaches to test security and privacy requirements of IoT cloud services from a user perspective. To allow for a mapping of tests to elements of the architecture of IoT cloud services, a generalized architecture for these services is derived. This test framework is then applied to some of the most popular IoT cloud services, i.e. *Amazon AWS IoT, Microsoft Azure* and *Google Cloud Platform*. In addition as an example for a smaller open source provider *ThingSpeak* is also considered.

The remainder of this paper is organized as follows: In Section 2 an overview about related work is summarized. Afterwards in Section 3 we present a generalized architecture of IoT cloud services and specify which elements of the architecture are specific for IoT cloud services. For these elements tests for security and privacy requirements are proposed in Section 4. The evaluation of the specified IoT cloud services based on the proposed test framework is described and discussed in Section 5. A summary and ideas for further research are stated in Section 6.

2 Related Work

IoT is an emerging field today, especially in connection with cloud approaches (see Botta et al. [9], [8]). This combination is especially important for smart city or mobility services, where IoT devices are distributed over a large geographical area. Important issues concerning IoT applications in general are the security and privacy challenges which are discussed in general by Zhao et al. [41], Zhang et al. [40], Abomara et al. [1], Jing et al. [24] and Ashraf et al. [4]. With a focus on the industrial internet of things, these challenges are investigated by Sadhegi et al. [34]. An investigation of security requirements and attacks based on an abstract IoT architecture are discussed in Hossain et al. [23]. The ENISA proposes threat taxonomies for IoT systems for different areas, see e.g. [5]. Vasilomanolakis et al. [38] state an in depth analysis of security and privacy requirements investigating beside network security, identity management and resilience which are typically considered also aspects of privacy and trust. Concerning the stated requirements various IoT reference architectures are assessed based on the respective specifications.

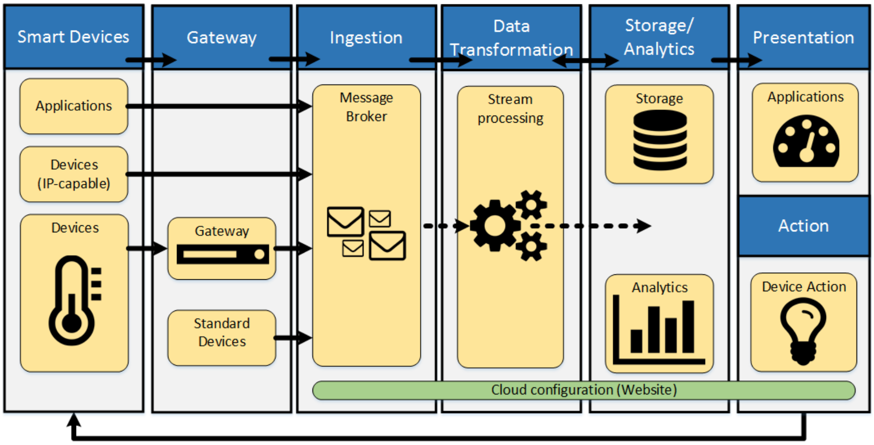


Figure 1: Generalized IoT cloud architecture

For cloud security in general Fernandes et al. [15] provide a comprehensive overview and propose a taxonomy for threats, vulnerabilities and attacks. For websites which are typically important elements of cloud services e.g. for dashboards and the configuration of services, lists of known and common web vulnerabilities are available [29]. Automated testing tools as e.g.[6] scan websites for typical vulnerabilities. Security aspects of the emerging IoT cloud services are investigated by Zhou et al. [42] with a focus on concepts for secure packet forwarding and privacy-preserving authentication. There a generic model is used as basis for the investigation. None of the existing cloud services in this area is considered.

An important security mechanism for cloud services in general is the use of firewalls, since public cloud services are reachable over the Internet. Ullrich et al. [37] propose security tests for firewalls of cyber-physical cloud computing, as an extension of earlier work focused on IaaS cloud firewalls in general [12]. In comparison to these investigations which are focused on network protocol vulnerabilities, the focus in this paper is enlarged to further elements of the IoT cloud architecture. Beside testing of firewalls of cyber-physical cloud services no test approaches for other elements of IoT cloud services are known.

3 Generalized IoT Cloud Architecture

IoT cloud services are emerging rapidly at the moment, see e.g. Microsoft Azure IoT (Figure 2), Google Cloud Platform (Figure 3), Amazon Web Services IoT, IBM and Oracle. The structure and components of these architectures have strong similarities. The basic idea is the concept of a data-flow, components as e.g. gateways, message broker and stream processing are present in all architectures. Based on these architectures we generalized an abstract architecture (Figure 1) which is used for further investigations.

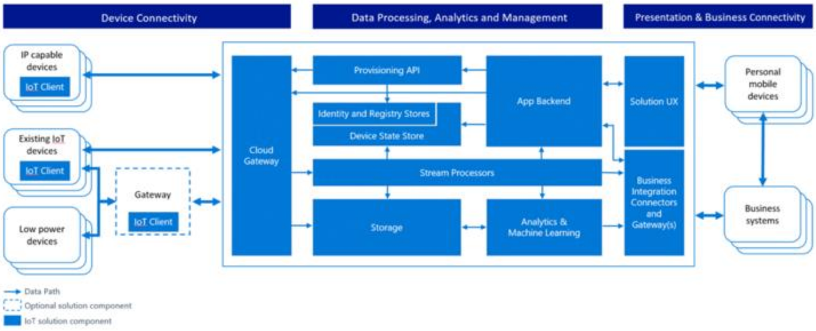


Figure 2: Microsoft Azure - IoT architecture [28]

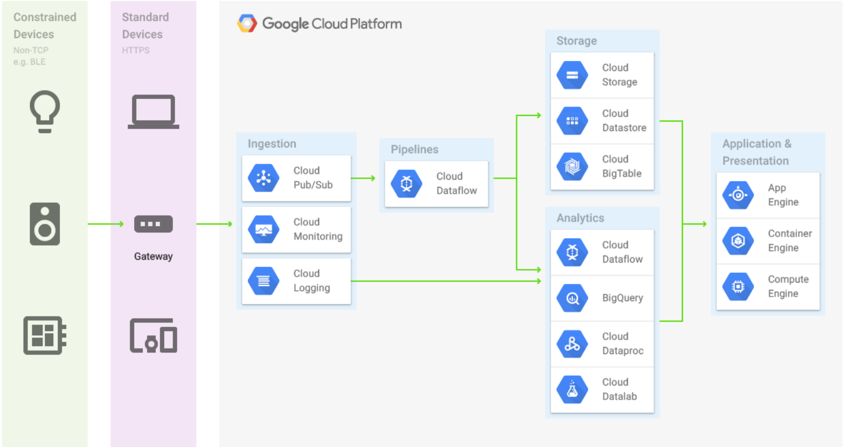


Figure 3: Google Cloud Platform - IoT data management [18]

3.1 Overview of IoT Cloud Architecture

The basis of IoT services are *smart devices*, typically equipped with sensors resp. actuators. Examples of sensors are temperature or light sensors, cameras, fingerprint-scanners, etc. Some of these devices are already IP-capable and are connected directly to the cloud. Devices and their data streams can also be integrated by specialized *gateway* systems that handle the cloud connection. Specialized lightweight operating systems provided by IoT cloud providers for *gateways* are Android Things [3] and Windows 10 IoT Core [27].

Data streams of devices are transfered to a central so-called *ingestor*. As communication protocols for these *message brokers* mainly Representational State Transfer (REST) based Web Services and Message Queue Telemetry Transport (MQTT) are used [25]. Afterwards the data streams are processed in the *data transformation* step. This step consists mainly of stream processing to transform, aggregate and/or enrich data. This preprocessed data stream can then be persisted. After these preprocessing steps the persisted data can be used in applications in the phase *storage/analytics*. Typical analytics steps are data mining or real time data analytics [36].

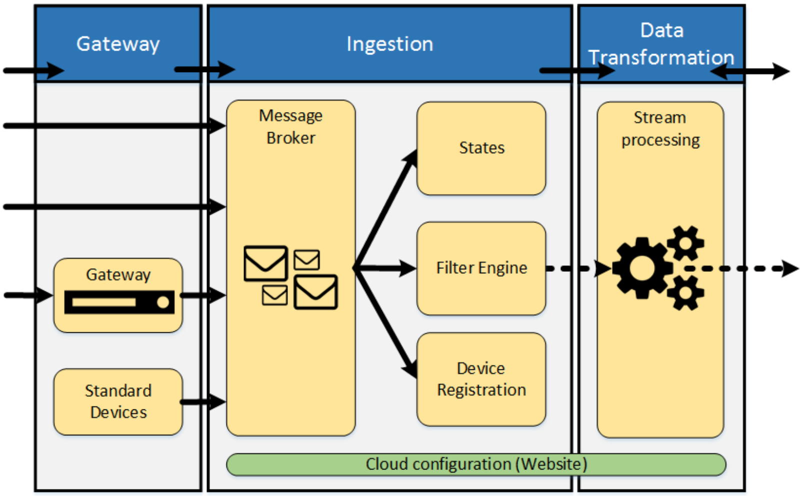


Figure 4: Structure of ingestor phase

Afterwards in the *presentation* phase data or derived results can be visualized e.g. on a dashboard. This visualization can either be provided by the cloud provider or could be realized as an external software which is connected through the cloud API. In addition devices with actuators can be remotely controlled based on this data (stated as *action* phase). Examples of actuators are e.g. switches in a smart home environment or smart traffic signs in the case of a smart city environment. The configuration of the IoT cloud service is realized by a corresponding web-interface.

3.2 Structure of Ingestor Phase

Figure 4 presents the typical structure of the *ingestor* phase which will be in the focus of the further investigation. All messages from the smart devices resp. gateways are sent to a central *message broker* which provides a communication interface. Usually RESTful Web services or MQTT brokers handle the inter-operable communication. The messages are typically based on a device-specific data model comprising device specific information and the current value of attributes like e.g. temperature and humidity. JavaScript Object Notation (JSON) is often used for these structured messages.

Every incoming message from the *message broker* is passed to several IoT specific elements. The *states* are a representation of the current values of the attributes of devices as specified by the data model. Incoming messages are updating the state of the corresponding device. Before a device can communicate with the cloud it has to be registered at the *device registration*. Every device gets a unique address while certificates/tokens ensure a secure authentication.

The *filter engine* is able to react directly on incoming messages passed from the *message broker*. A filter rule consists of a condition and a resulting action in case the condition matches. If for example a temperature sensor sends a new temperature value, the *filter engine* checks if the value is above a threshold to start the air-conditioner. For further processing the resulting data stream is then transfered to other cloud service in the following phases.

4 Security Testing of IoT Clouds

To ensure security and privacy requirements all elements of IoT cloud services should be tested in a holistic way. For users of these cloud services this is not possible because of lack of transparency. We investigate in the following which elements can be tested from "outside". Technically oriented tests with a strong potential for automation are proposed for the IoT specific elements (Figure 4).

4.1 Beyond Security Testing

Some of the elements of the general architecture are not specific for IoT cloud services. The standard cloud services for (*stream processing, data transformation, storage, analytics*) are extensively covered by [15]. These services could not be transparently tested from outside. As an example for vulnerabilities, Jones [31] showed how to exploit AWS Lambda, a stream processing service, by abusing undocumented features.

The only possibility to address security and privacy requirements concerning elements of public cloud services that are not transparent are certifications based on standardized audits (cf. Sunyaev et al. [35]). The European Union Agency for Network and Information Security (ENISA) [13] developed based on the *European Cloud Strategy* [11] a *Cloud Certification Schemes Metaframework (CCSM)* and a derived *Cloud Certification Schemes List (CCSL)* where certifications for cloud services are listed according to the specified criteria. The *National Institute of Standards and Technology (NIST)* [22] defined a cloud reference architecture, use cases of the public sector and investigated existing standards based on the requirements of federal organizations. The most important standards for security and privacy in cloud computing are ISO/IEC 27001:2013 where requirements for an information security management system are defined, ISO/IEC 27017:2015 stating information security controls for cloud services and ISO/IEC 27018:2014 about protection of personally identifiable data in public clouds.

The *smart devices* itself are not part of the IoT cloud services. Security issues for smart devices are extensively covered by [1, 34, 40, 41]. Because of the characteristics of these devices, i.e. according to [38] "the uncontrolled environment, the heterogeneity, the need for scalability and the constrained resources", instead of defining generalized test cases, testbeds with a focus on certain use cases are proposed, see e.g. Hahn et al. [20].

4.2 Security and Privacy Testing

The focus of security and privacy tests presented here are the specialized IoT cloud elements of the *ingestor* phase (i.e. message broker, states, device registration, filter engine) and the network communication between *gateways* resp. *devices* and *message broker*. In the following we describe the investigated tests grouped according to corresponding architecture elements. An overview is given in Table 1.

Table 1: Implemented Test Cases

ID	*Name*
A	**Network**
1	Open Ports
2	White-Listing
3	ICMP Services
4	Message Integrity
5	Message Confidentiality
6	Mutual Authentication
B	**Message Broker**
7	Oversize Payload
8	Message Throughput
9	Parallel Connections
10	Resource Information
11	Authentication
C	**States**
12	Message Fuzzing
13	Updating States
D	**Device Registration**
14	Authentication Manager
15	Certificates
16	Revoke certificate/token
E	**Filter Engine**
17	Rule Consistency
18	Rule Injection
19	Rule Typecheck
F	**Compliance**
20	Security Standards
21	Geographic Areas
22	Privacy Regulation

A: Network

In the network configuration, the number of *open ports* should be restricted to the absolute minimum and all open ports should be documented. To avoid connections from unknown devices already at the network layer, it should be possible to configure that only trusted devices with known IP addresses are allowed to contact the *message broker*. This *white-*

listing can be tested by sending IP packets from arbitrary IP addresses. These connections should be blocked also if the device has a valid certificate resp. token.

Furthermore Internet Control Message Protocol (ICMP) services as ping and traceroute should not be provided since they allow an attacker to gain information about the network configuration. To prevent eavesdropping, interception and manipulation of network packets, message authentication codes (MAC) and encryption of messages should be standard security measurements ensuring the integrity and confidentiality of messages. Furthermore it is tested if the cloud provider offers mutual authentication to ensure that both client and server are connected to the right communication partner.

B: Message Broker

The behavior of the *message broker* regarding overloading with messages and clients is examined. The behavior is tested by sending messages whose length exceeds the documented limits. Another test takes a look at the maximum message throughput of the server and if this is restricted. Also the *message broker* is supposed to limit the maximum number of parallel connections. Furthermore it is tested if information about the existence of a project is publicly available (e.g. by revealing information about a project specific domain). Also it is tested that a connection to the *message broker* can only be established after successful authentication of the device resp. gateway.

C: States

As already described, *states* are a representation of the current attributes of the device. One test uses fuzzing to generate random invalid messages to detect parsing errors. Another test is based on the message format and tests if the nesting depth of structured data formats e.g. JSON is restricted.

D: Device Registration

Each device is supposed to be registered separately. Therefore an authentication manager is necessary where devices need to be registered. Furthermore providers should offer certificates for securing the communication. Also the behavior is tested when revoking a certificate while the corresponding device is still connected.

E: Filter Engine

The first test in this group examines the consistency of filter rules and what opportunities for the assignment of permissions and monitoring of modifications exist. A further test tries to “bypass” the *filter engine* similar to SQL injections (e.g. temperature=”100” or 1=1). Furthermore it is evaluated if filters are able to deal correctly with different data types.

F: Compliance

Since several elements of IoT cloud services cannot be transparently tested, it is important also to consider audits based on standards and security and privacy related certifications. From a compliance point of view it is important in which country data is stored and processed, since based on legal regulations government agencies might have lawful access to data and there are legal restrictions concerning the transfer of personal data to third countries. Concerning the processing of personal data also legal regulations need to be considered. In Europe processing of personal data has to be compliant with the Data Protection Directive (95/46/EC) \cite{commission1995dpd} resp. national legislation. From May 25th, 2018 the General Data Protection Regulation (Regulation (EU) 2016/679) [30] will be set into force.

5 Results of the Study and Discussion of the Results

The test cases described in Section 4 were applied to the IoT cloud providers *Amazon AWS IoT*, *Microsoft Azure*, *Google Cloud Platform* and *ThingSpeak* in July 2017. Table 2 summarizes the results. A test can either be fulfilled, partially fulfilled (i.e. with limitations) or not fullfilled. In the following mainly the results of all tests which are either partially fulfilled or not fulfilled are explained and discussed.

5.1 Test Results

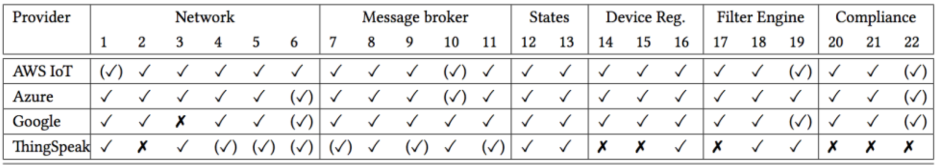
A: Network

AWS IoT allows port 8192 which is not specified in their documentation [2]. White listing is supported by all providers except ThingSpeak. Google is the only provider allowing ICMP services (ping/traceroute). Amazon, Microsoft and Google encrypt network communication by default and use message authentication codes. ThingSpeak is the only provider offering both, secure and non-secure communication (test cases 4, 5). Mutual authentication is only supported by AWS IoT. They are offering their own root certificate. Azure and ThingSpeak are not supporting any kind of mutual authentication except the TLS handshake. Google proposes “automatic mutual authentication” [17].

B: Message Broker

When sending a message to the ThingSpeak service that exceeds the maximum packet length (test case 7) the web server reveals its name plus version number which is outdated.

Table 2: Results of test cases per cloud provider: ✓test fulfilled, (✓) partially fulfilled, ✗not fulfilled



Provider	Network						Message broker					States		Device Reg.			Filter Engine			Compliance		
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
AWS IoT	(✓)	✓	✓	✓	✓	✓	✓	✓	✓	(✓)	✓	✓	✓	✓	✓	✓	✓	✓	(✓)	✓	✓	(✓)
Azure	✓	✓	✓	✓	✓	(✓)	✓	✓	✓	(✓)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	(✓)
Google	✓	✓	✗	✓	✓	(✓)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	(✓)	✓	✓	(✓)
ThingSpeak	✓	✗	✓	(✓)	(✓)	(✓)	(✓)	✓	(✓)	✓	(✓)	✓	✓	✗	✗	✓	✗	✓	✓	✗	✗	✗

ThingSpeak is able to run with multiple MQTT connections to publish messages. Subscribing to topics in MQTT is not supported. AWS and Azure are providing unique URLs for each of their IoT projects. The other providers are using a centralized communication server for all of their projects (test case 10). ThingSpeak is using API keys for the authentication. When posting on their RESTful API, only the API key is necessary, no more information as e.g. the project ID is needed. Therefore brute-force attacks could be possible.

C: States

The structure of the data models of all providers is well defined. Limits like the nesting depth in JSON are stated. Malformed messages are discarded.

D: Device Registration

In the ThingSpeak cloud, devices could not explicitly be registered. Furthermore it is not possible to use certificates. There is one API key which is used for all devices. Revocation of certificates or API keys is possible for all providers (test case 16).

E: Filter Engine

ThingSpeak does not have a role and rights management system. One account is used per project whose rights can not be limited. Concerning test case 18, rule injection, for none of the cloud services it was possible to "bypass" the filter engine with malformed statements. AWS and Google do not have a type check mechanism for the filter engine. Hence it is not possible to check if values are in the right format resp. data-type.

F: Compliance

Microsoft, Amazon and Google present a comprehensive amount of certifications including the standards stated in Section 4 ISO/IEC 27001, ISO/IEC 27017 and ISO/IEC 27018. These providers support the restriction of data processing to specific geographic regions and state compliance with European Privacy Regulation, i.e. the Data Protection Directive (95/46/EC) [14] and the General Data Protection Regulation (Regulation (EU) 2016/679). Thingspeak does not address these compliance issues. Since the EU-U.S. Privacy Shield which is the successor of the Safe Harbour agreement is being questioned at the moment [7], "partial fulfilled" is marked in Table 2 concerning privacy regulations because of the unclear political situation. The risk needs to be evaluated by users based on the use case and data to be processed. Additional measurements as e.g. privacy enhancing technologies (PET) (see Hoepman [21]) as anonymization or pseudonymization should be considered.

5.2 Discussion of Test Results

The results of the tests show in the first instance the focus of the IoT cloud providers: The aim of AWS, Azure and Google is to built platforms consisting of the IoT cloud architecture and accompanying lightweight operating systems for the realization of gateways, such that other service providers realize their IoT services based on the platforms. In contrast ThingSpeak, which is described as open IoT service, addresses merely private users and their projects. Hence the focus of ThingSpeak is to make the usage of their platform as easy as possible without an explicit focus on security and privacy. Whereas the other providers state profound security concepts, provide a broad range of certifications addressing security, privacy and other compliance issues. Considering the groups of security and privacy requirements as proposed by Vasilomanolakis et al. [38], network security, identity management and resilience are adequately addressed by AWS, Google and Azure. Concerning privacy and trust, only attestations by certifications are present, since the cloud platforms are not transparent. Additional technical measurements as privacy enhancing technologies are not in the focus of the providers. The resilience of cloud services could be further enhanced by developing distributed cloud architectures as in intercloud approaches.

6 Conclusion

In this paper security and privacy of IoT cloud services were investigated, based on a generalized IoT cloud architecture derived from existing cloud services. Based on this architecture test cases concerning security and privacy were proposed and evaluated for four IoT cloud services. The IoT cloud services of AWS, Google and Azure show an adequate level of security to be used as a platform for IoT cloud services. Privacy and trust requirements are mainly addressed by certifications. Technical measurements as privacy enhancing technologies are not provided. Future work will address the investigation of further elements of the architecture as e.g. the stream processing engine and to develop further test cases especially concerning privacy and trust.

Acknowledgments

This work was supported by the Ministry for Science and Culture of Lower Saxony as part of SecuRIn (VWZN3224).

References

[1] Mohamed Abomhara and Geir M Køien. 2014. Security and privacy in the Internet of ings: Current status and open issues. In Privacy and Security in Mobile Systems (PRISMS), 2014 International Conference on. IEEE, 1–8.

[2] Amazon Web Services, Inc. 2017. Protocols. (March 2017). http://docs.aws. amazon.com/iot/latest/developerguide/protocols.html

[3] Android Things. 2017. Android Things. (Feb. 2017). https://developer.android. com/things

[4] Qazi Mamoon Ashraf and Mohamed Hadi Habaebi. 2015. Autonomic schemes for threat mitigation in Internet of Things. Journal of Network and Computer Applications 49 (2015), 112–127.

[5] D Barnard-Wills, L Marinos, and S Portesi. 2014. Threat landscape and good practice guide for smart home and converged media. enisa, Tech. Rep. (2014).

[6] Beyond Security. 2017. ScanMyServer: Test the security of your website, web server or blog - Free! (Feb. 2017). https://www.scanmyserver.com/

[7] Stephanie Bodoni. 2017. If Trump Spoils Privacy Pact, We'll Pull It, EU Official Warns. Bloomberg (March 2017). https://www.bloomberg.com/news/articles/ 2017- 03- 02/if-trump- spoils- privacy-pact-we-ll-pull-it-eu-official- warns.

[8] Alessio Botta, Walter De Donato, Valerio Persico, and Antonio Pescapé. 2016. Integration of cloud computing and internet of things: a survey. Future Generation Computer Systems 56 (2016), 684–700.

[9] A.Botta, W. de Donato, V. Persico, and A. Pescap. 2014. On the Integration of Cloud Computing and Internet of Things. In 2014 International Conference on Future Internet of Things and Cloud. 23–30. DOI: http://dx.doi.org/10.1109/ FiCloud.2014.14

[10] Cesar Cerrudo. 2015. An emerging us (and world) threat: Cities wide open to cyber attacks. Securing Smart Cities (2015). http://www.ioactive.com/pdfs/IOActiveHackingCitiesPaperCesarCerrudo.pdf.

[11] European Union Communication. 2012. Communication from the Commission to the Council, the European Parliament, the European Economic and Social Committee and the Committee of the Regions: Unleashing the Potential of Cloud Computing in Europe. (2012). http://eur-lex.europa.eu/legal-content/EN/TXT/ ?uri=CELEX%3A52012DC0529.

[12] Jordan Cropper, Johanna Ullrich, Peter Frühwirt, and Edgar Weippl. 2015. The role and security of firewalls in iaas cloud computing. In Availability, Reliability and Security (ARES), 2015 10th International Conference on. IEEE, 70–79.

[13] M.A.C. Dekker and D. Liveri. 2014. CCSM - Cloud Certification Schemes Metaframework. ENISA, Tech. Rep. (2014). https://resilience.enisa.europa.eu/cloud-computing-certification/cloud-certification-schemes-metaframework.

[14] European Union Directive. 1995. Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. (1995). http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX: 31995L0046:EN:HTML.

[15] Diogo AB Fernandes, Liliana FB Soares, João V Gomes, Mário M Freire, and Pedro RM Inácio. 2014. Security issues in cloud environments: a survey. International Journal of Information Security 13, 2 (2014), 113–170.

[16] Thomas Fox-Brewster. 2017. Ransomware Crooks Demand 70,000$ After Hacking San Francisco Transport System – UPDATED. (Feb. 2017). https://www.forbes.com/sites/thomasbrewster/2016/11/28/san-francisco-muni-hacked-ransomware/#2cc9c7ec4706

[17] Google. 2017. Google Infrastructure Security Design Overview. (March 2017). https://cloud.google.com/security/security- design/

[18] Google. 2017. Overview of Internet of things. (July 2017). https://cloud.google. com/solutions/iot- overview

[19] Jayavardhana Gubbi, Rajkumar Buyya, Slaven Marusic, and Marimuthu Palaniswami. 2013. Internet of Things (IoT): A Vision, Architectural Elements, and Future Directions. Future Gener. Comput. Syst. 29, 7 (Sept. 2013), 1645–1660.

[20] Adam Hahn, Aditya Ashok, Siddharth Sridhar, and Manimaran Govindarasu. 2013. Cyber-physical security testbeds: Architecture, application, and evaluation for smart grid. IEEE Transactions on Smart Grid 4, 2 (2013), 847–855.

[21] Jaap-Henk Hoepman. 2014. Privacy design strategies. In IFIP International Infor- mation Security Conference. Springer, 446–459.

[22] Michael Hogan and Sophie Sokol. 2014. NIST Cloud Computing Standards Roadmap. NIST Special Publication 500-291, Version 2 (2014). https://www.nist.gov/sites/default/les/documents/itl/cloud/NISTSP-500-291 Version-22013June18FINAL.pdf.

[23] Md Mahmud Hossain, Maziar Fotouhi, and Ragib Hasan. 2015. Towards an analysis of security issues, challenges, and open problems in the internet of things. In Services (SERVICES), 2015 IEEE World Congress on. IEEE, 21–28.

[24] Qi Jing, Athanasios V Vasilakos, Jiafu Wan, Jingwei Lu, and Dechao Qiu. 2014. Security of the Internet of Things: perspectives and challenges. Wireless Networks 20, 8 (2014), 2481–2501.

[25] Vasileios Karagiannis, Periklis Chatzimisios, Francisco Vazquez-Gallego, and Jesus Alonso-Zarate. 2015. A survey on application layer protocols for the internet of things. Transaction on IoT and Cloud Computing 3, 1 (2015), 11–17.

[26] Brian Krebs. 2016. Hacked Cameras, DVRs Powered Today s Massive Internet Outage. (Oct. 2016). https://krebsonsecurity.com/2016/10/ hacked- cameras- dvrs- powered-todays- massive- internet- outage/

[27] Microsoft. 2017. Windows 10 IoT Core. (Feb. 2017). https://developer.microsoft . com/en-us/windows/iot

[28] Microsoft Azure. 2017. IoT Security Architecture. (July 2017). https://docs. microsoft .com/en- us/azure/iot- suite/iot- security- architecture

[29] Top OWASP. 2013. 10: Ten Most Critical Web Application Security Risks. (2013).

[30] European Union Regulation. 2016. 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). (2016). http://data.europa.eu/eli/reg/2016/679/oj.

[31] Rich Jones. 2017. Gone in 60 Milliseconds -Intrusion and Exfiltration in Server- less Architectures. (March 2017). https://media.ccc.de/v/33c3-7865-gone-in-60-milliseconds

[32] Eyal Ronen, Colin O Flynn, Adi Shamir, and Achi-Or Weingarten. 2016. IoT Goes Nuclear: Creating a ZigBee Chain Reaction. (2016).

[33] Ronen, Eyal and O Flynn, Colin and Shamir, Adi and Weingarten, Achi-Or. 2017. IoT Goes Nuclear: Creating a ZigBee Chain Reaction. (Feb. 2017). http: //iotworm.eyalro.net/

[34] Ahmad-Reza Sadeghi, Christian Wachsmann, and Michael Waidner. 2015. Secu- rity and privacy challenges in industrial internet of things. In Design Automation Conference (DAC), 2015 52nd ACM/EDAC/IEEE. IEEE, 1–6.

[35] Ali Sunyaev and Stephan Schneider. 2013. Cloud services certification. Commun. ACM 56, 2 (2013), 33–36.

[36] Ralf Tonjes, P Barnaghi, M Ali, A Mileo, M Hauswirth, F Ganz, S Ganea, B Kjærgaard, D Kuemper, Septimiu Nechifor, and others. 2014. Real time iot stream processing and large-scale data analytics for smart city applications. In poster session, European Conference on Networks and Communications.

[37] Johanna Ullrich, Jordan Cropper, Peter Fru ̇hwirt, and Edgar Weippl. 2016. The role and security of firewalls in cyber-physical cloud computing. EURASIP Journal on Information Security 2016, 1 (2016), 18.

[38] Emmanouil Vasilomanolakis, Jo ̈rg Daubert, Manisha Luthra, Vangelis Gazis, Alex Wiesmaier, and Panayotis Kikiras. 2015. On the Security and Privacy of Internet of Things

Architectures and Systems. In Secure Internet of ings (SIoT), 2015 International Workshop on. IEEE, 49–57.

[39] Kyle York. 2016. Dyn Statement on 10/21/2016 DDoS A ack. (Oct. 2016). http://hub.dyn.com/static/hub.dyn.com/dyn- blog/ dyn- statement- on- 10- 21-2016-ddos-attack.html

[40] Zhi-Kai Zhang, Michael Cheng Yi Cho, Chia-Wei Wang, Chia-Wei Hsu, Chong-Kuan Chen, and Shiuhpyng Shieh. 2014. IoT security: ongoing challenges and research opportunities. In Service-Oriented Computing and Applications (SOCA), 2014 IEEE 7th International Conference on. IEEE, 230–234.

[41] Kai Zhao and Lina Ge. 2013. A survey on the internet of things security. In Computational Intelligence and Security (CIS), 2013 9th International Conference on. IEEE, 663–667.

[42] Jun Zhou, Zhenfu Cao, Xiaolei Dong, and Athanasios V Vasilakos. 2017. Security and Privacy for Cloud-Based IoT: Challenges. IEEE Communications Magazine 55, 1 (2017), 26–33.

Why Did I End Up Living in a Cave? Risks of IoT at Home

NEREA SAINZ DE LA MAZA DOÑABEITIA, MIGUEL HERNÁNDEZ BOZA,
JAVIER JIMÉNEZ DEL PESO & JOSÉ IGNACIO ESCRIBANO PABLOS

Abstract Technology revolution is unstoppable, we live in a world in which every device is connected to the Internet. The Internet of Things is boosting due to the reduction of size and price of the sensors.

At home, we can find dozens of gadgets from smart cars that can park themselves to fridges which know what kind of products are being stored inside and determine whenever a food item needs to be replenished.
However, the problem is that many of these devices are being made the same way as 20 years ago without focusing on the security issues.

In this paper, we will research the vulnerabilities and security problems that different devices may have and we propose a model to estimate the risk score of them. Some of the variables to calculate the score are authentication type, default credentials, vulnerabilities, hacking news and number of exposed devices to the Internet (more than
20\,000 discovered). Then, we combine these scores to find out how secure your home is, based on their individual values.

This model is used by a web application that lets the users select the IoT devices they have at home and returns the security risk of suffering an attack.

Keywords: • IoT • smart homes • cybersecurity • risk assessment • vulnerabilities • privacy•

CORRESPONDENCE ADDRESS: Nerea Sainz de la Maza Doñabeitia, Innovation 4 Security, Avenida de Burgos 16D, 28036 Madrid, Spain, e-mail: nerea.sainz@i4s.com. Miguel Hernández Boza, Innovation 4 Security, Avenida de Burgos 16D, 28036 Madrid, Spain, e-mail: miguel.hernandez@i4s.com. Javier Jiménez del Peso, Innovation 4 Security, Avenida de Burgos 16D, 28036 Madrid, Spain, e-mail: javier.jimenezdelpeso@i4s.com. José Ignacio Escribano Pablos, Innovation 4 Security, Avenida de Burgos 16D, 28036 Madrid, Spain, e-mail: joseignacio.escribano@i4s.com.

DOI https://doi.org/10.18690/978-961-286-114-8.11 ISBN 978-961-286-114-8

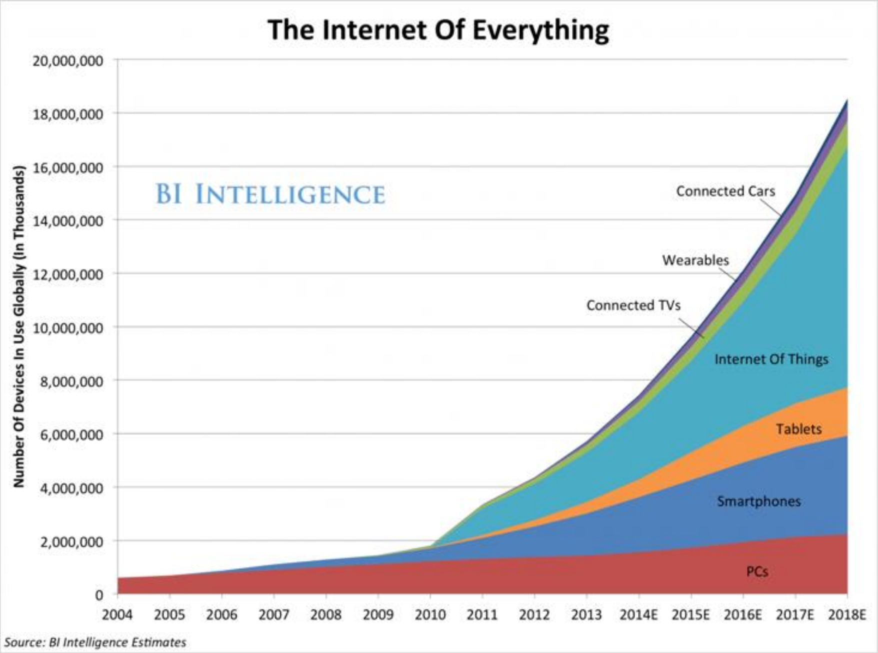
Available at: http://press.um.si.

1 Introduction

The Internet of Things (IoT) is a global infrastructure for the information society, enabling advanced services by interconnecting (physical and virtual) things based on existing and evolving interoperable information and communication technologies [26].

The number of IoT devices is continuously increasing. It will grow to 26 billion in 2020, according to Gartner [23]. It will be necessary to use IPv6 instead of IPv4 in order to support the large number of active devices, addresses will be insufficient.

However, from the cybersecurity point of view, the cost reduction means that the devices may not implement enough security procedures. This provokes huge security holes, leakage of sensitive data or Denial of Service (DoS) if the device is encrypted by using specific *ransomware* [29]. But what would happen if these devices controlled your home access?



Source: BI Intelligence Estimates

Figure 1: The Internet of Everything evolution [25]

We analysed dozens of IoT devices after an exhaustive research in cybersecurity webpages, CVE's entries and Shodan search engine that are presented in the following sections. Then we will estimate the risk score of the house depending on the devices it contains.

The paper is divided into the following sections: the research highlights can be found in Section 2, in Section 3 we explain the proposed model, in Section 4 we introduce the developed web application, and finally, in Section 5 we present the conclusions.

2 Analyzed devices: highlights

In this section we present the highlights from the investigation. A more detailed analysis can be found on Table 3:

- Lighting control system: **HUE Personal Wireless Lighting** from Philips [24] has a vulnerability (CVE-2014-4883 [16]) that affects the integrity and confidentiality of communications. It was resolved in December 2014 and it is needed to update the devices via HUE app. HUE uses HTTP making easy to attack it via *Man-in-the-middle* allowing to control it without any kind of authentication. An example can be found here [1].
- Blinds: **Loxone Blind Control** [28] has default credentials *admin:admin*. We found around 200 devices exposed[1] on the Internet. In case of the default credentials have not been changed it may be a security problem allowing full access to the platform.
- Televisions: Maybe the most famous case is the use of **Samsung Smart TVs** by CIA agents to spy capturing audio recordings. They can also be used for recovering the Wi-Fi keys and accessing any usernames and passwords stored on the TV browser [44]. Other examples are the use of **Philips SmartTV** to steal user cookies and other sensitive data [35] or ransomware on **LG Smart TV** [36].

  There are some vulnerabilities related to smart TVs that allow DoS attacks and arbitrary code execution:

  CVE-2015-8040 [20].
  CVE-2015-8039 [19].
  CVE-2014-9266 [17].
  CVE-2012-4330 [11].
  CVE-2012-2210 [10].

  We found near 12 000 Samsung devices exposed on the Internet. We also found exposed TVs from LG and Sony, but fewer devices than Samsung TVs.
- Dishwasher: **Miele Washer disinfector** [32] used in hospitals has a traversal vulnerability (CVE-2017-7240 [21]) which allows access to directories and information that can be exploited to infect the machine with malware [45].

Fridges: the fridge **Samsung RF28HMELBSR** does not validate SSL certificates to secure the Gmail integration and allows access to the network and steal the Gmail credentials [22].

Toys: **Smart Toy Bear** left $800 000$ customer credentials and 2 million audio recordings exposed [43].
Sexual toys: the **Siime Eye** [39], a dildo with a tiny camera, that allows the user to stream videos. The Siime Eye creates a Wi-Fi Access Point whose default password is “88888888”. That way, anyone in range can connect to it, access the login portal, where the user is “admin”' and the password is blank, and watch the live streaming [38].

Webcams: we analysed 3 models: **SQ-Webcam** [40] **Webcamxp** [46] and **YawCam** [48] and we found 196, 1146 and 1317 devices exposed, respectively. They have some XSS (Cross-Site Scripting) vulnerabilities:

- CVE-2004-2094 [4].
- CVE-2005-1189 [5].
- CVE-2005-1190 [6].
- CVE-2008-5674 [7].
- CVE-2008-5862 [8].
- CVE-2003-1479 [3].

Printers: we analysed 2 models: **PapercutMF** [34] and **Toshiba TopAccess** [41]. PapercutMF has 3 vulnerabilities that allow DoS attacks and *Cross-Site Request Forgery* (CSRF):

- CVE-2014-2659 [15].
- CVE-2014-2658 [14].
- CVE-2014-2657 [13].

TopAccess has 2 CSRF vulnerabilities:

- CVE-2012-1239 [9].
- CVE-2014-1990 [12].

Cars: Charlie Miller and Chris Valasek exposed the security vulnerabilities in automobiles by hacking a **Jeep Cherokee** remotely [47]. The documented vulnerability is CVE-2015-5611 [18].

2.1 Exposed devices

On Table 1 it is shown the number of exposed devices on the Internet using Shodan [37] divided in 4 rooms

- Living Room includes lighting system, blinds, alarm system, TV, speaker, etc.
- Kitchen includes fridge, washing machine, microwave, coffeemaker, smoke detector, etc.
- Bedroom includes toys, wearables, webcams, printers, home control, etc.
- Garage/Garden includes irrigation system, cars, garage doors, etc.

Table 1: Number of exposed devices in each room.

Room	Found devices
Living Room	20 409
Kitchen	26
Bedroom	3 633
Garage/Garden	75
Total	**24 143**

3 Risk assessment model

Several risk assessment models have been published. Diego Mendez et al. [31] analyses papers published for the IoT, focusing on IoT vulnerabilities as well as the security challenges in the data confidentiality, integrity, availability and privacy. Liu et al. [27] proposed a dynamic risk assessment methodology for the IoT inspired by the artificial immune system. M. Mohsin et al. [33] proposed a framework that generates threat models utilized to compute the likelihood and attacker cost for exploiting IoT vulnerabilities. Yair Meidan et al. [30] used machine learning techniques such as Random Forest to analyse network traffic data in order to detect unauthorized IoT devices.

We propose a model to estimate the potential risk of a house with n IoT devices. The risk of the house, R_v, is a value between 0 and 1 where a value close to 0 means that our home is secure and a value very close to 1 means important security issues. It is calculated as a weighted arithmetic mean of the individual devices scores, following the Equation 2,

$$R_v = \frac{\sum_{i=1}^{n} \omega_i s_i}{\sum_{i=1}^{n} \omega_i} \tag{1}$$

where ω_i is the criticality level (see Section 3.1) from device i, s_i is the *score* from device i and n is the number of devices.

3.1 Criticality level

Not all the devices suppose the same security risk for the user. We will assign different weights depending on the impact on the users:

- 0.4, if it affects physical access.
- 0.3, if it affects privacy.
- 0.2, if it affects home integrity.
- 0.1, otherwise.

The Table 2 shows the criticality values depending on the device class.

Table 2: Criticality values for each type of device

Device	Value	Device	Value
Lightning system	0.1	Washing machine	0.2
Blind	0.1	Dishwasher	0.2
Temperature system	0.2	Microwave	0.2
Alarm system	0.4	Coffeemaker	0.1
Lock	0.4	Smoke detector	0.4
TV	0.3	Garden control	0.1
Paper bin	0.1	Lift	0.4
Toy	0.3	Car	0.4
Webcam	0.4	Garage doors	0.4
Speaker	0.1	Hard disk	0.3
Cleaning robot	0.2	Printer	0.2
Fridge	0.2	Home Control	0.4

3.2 Score

After the investigation done in Section 2, we selected 8 variables to estimate the device score, s_i. The values that the variables can take:

- Authentication:

$$v_0 = \begin{cases} 0, & \text{with authentication} \\ 1, & \text{without authentication} \end{cases}$$

- Authentication type:

$$v_1 = \begin{cases} 1, & \text{on the device} \\ \frac{1}{2}, & \text{using OAuthh 2.0} \end{cases}$$

- HTTP:

$$v_2 = \begin{cases} 0, & \text{if it uses HTTPS} \\ 1, & \text{if it uses HTTP} \end{cases}$$

- Default credentials:

$$v_3 = \begin{cases} 0, & \text{if it has default credentials} \\ 1, & \text{if it has not default credentials} \end{cases}$$

- Devices exposed:

$$v_4 = \begin{cases} 0, & \text{if devices exposed were not found} \\ 1, & \text{if devices exposed were found} \end{cases}$$

- CVE's:

$$v_5 = \begin{cases} 0, & \text{if there is not any CVE related} \\ 1, & \text{if there is some CVE related} \end{cases}$$

- News:

$$v_6 = \begin{cases} 0, & \text{if there is not any vulnerability news} \\ 1, & \text{if there is some vulnerability news} \end{cases}$$

- Technical documentation:

$$v_7 = \begin{cases} 0, & \text{if there is not any technical doc} \\ 1, & \text{if there is some technical doc} \end{cases}$$

We assign different weights to the variables depending on their importance after the research. The score from each device is calculate as

$$s = 0.25v_0 + 0.1v_1 + 0.12v_2 + 0.1v_3 + 0.15v_4 + 0.18v_5 + 0.05v_6 + 0.05v_7$$

In the application we use the percentage instead, i.e. the result of multiplying the risk score by 100. On Table 3 it is shown the percentage associated to the different devices evaluated in Section 2.

4 Web application

We developed a web application which includes all the data previously presented and automatically calculates the potential risk score of a home based on the input that the user configures using the model described on Section 3.

4.1 Architecture

The application has been implemented using the MEAN stack architecture [2].

Backend. The backend allows to save the analysed data devices together with the score given by the Equation 2 in a MongoDB, using a REST API for querying.

Frontend. The frontend is responsible for loading and displaying the information given by the API and estimating the total risk of a house using the model explained in Section 3.

4.1.1 Views

The application consists of three views[2]:

- Devices catalogue (Figure 2): here is a sample of the IoT devices that can be found in a house. It is possible to search by name, room location or type. The user can add devices to estimate the risk of the house.

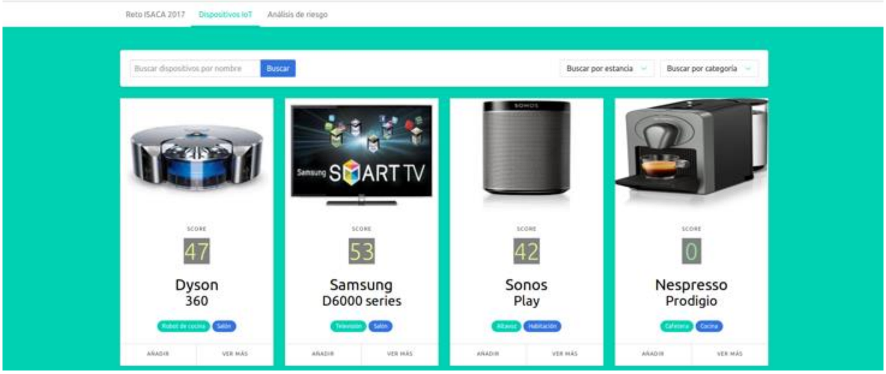
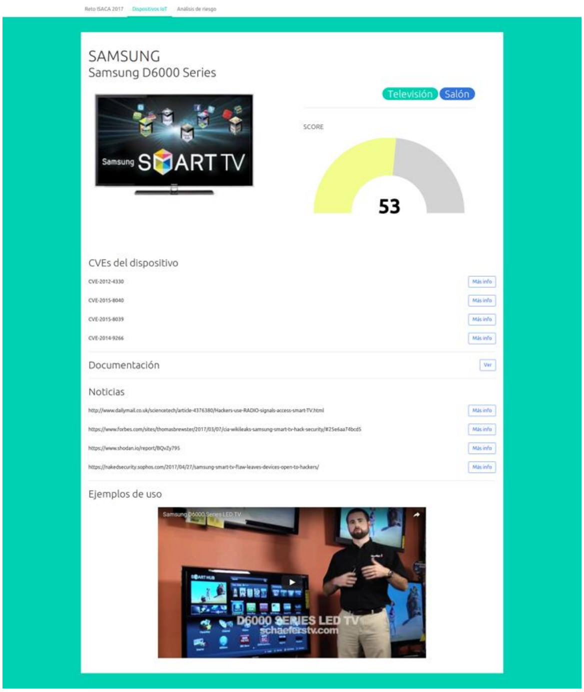


Figure 2: Devices catalogue.

- Device details view (Figure 3): if a device is clicked, a more detailed view of the product is shown: the score, CVEs, documentation, hacking news, related videos, etc.
- Risk score view (an example is given in Figure 3): once we select all the devices we have at home, in this view we can check the final risk score and find out how secure is our home.



SAMSUNG
Samsung D6000 Series

CVEs del dispositivo

CVE-2012-4330
CVE-2015-8040
CVE-2015-8039
CVE-2014-9266

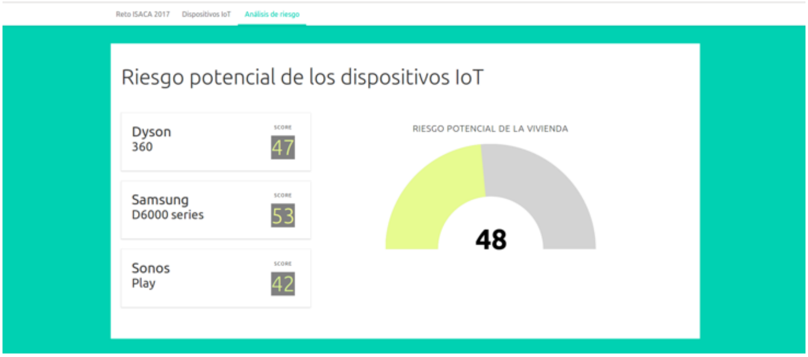
Documentación

Noticias

http://www.dailymail.co.uk/sciencetech/article-4376380/Hackers-use-RADIO-signals-access-smart-TV.html
https://www.forbes.com/sites/thomasbrewster/2017/03/07/cia-wikileaks-samsung-smart-tv-hack-security/#25e6aa74bcd5
https://www.shodan.io/report/BQvZy795
https://nakedsecurity.sophos.com/2017/04/27/samsung-smart-tv-flaw-leaves-devices-open-to-hackers/

Ejemplos de uso

Figure 3: Devices details view.



Riesgo potencial de los dispositivos IoT

Dyson 360	47
Samsung D6000 series	53
Sonos Play	42

Figure 4: Risk score example calculated with three devices: a cleaning robot with a scoring of 47, a TV with 53 and a couple of speakers with 42 (the individual scoring from each device is extracted from Table 3. Using a criticality Using a criticality value of 0.1 for the speakers, 0.2 for the cleaning robot and 0.3 for the TV and then we apply Equation 1. We obtain a percentage home risk equal to 48, meaning that our house is only "half" secure and we may have some insecure configuration in any of our IoT devices.

5 Conclusions and future work

Through this research we realise that the manufacturers do not use any standard when developing IoT devices and consequently, there is not homogenity between them. It is needed a colaboration between companies to implement a common security system.

It is easy to take control of a large variety of devices without a high hacking knowledge, because in many cases there is not any kind of authentication or default credentials are used. And as we saw on previous sections, many of them are exposed on the Internet and anyone can access them.

Moreover, there is a lack of awareness by the final users of the IoT gadgets about the security issues that the producers do not include in the user manual.

Taking this into account, we developed a web application to rise awareness of users and companies about the security that all the devices should implement to make a safer home. It can also be helpful for insurance companies to estimate the risk score of their customers' houses.

The IoT developers should be the responsibles for solving this insecurity problems and create some standards to improve security during the fabrication process instead of updating with patches. If the society continues ignoring this issue, the news about hacked microwaves, fridges or TVs will increase avoiding the opportunities that the Internet of Things offer in terms of quality of life.

For further research, it is important to create new approaches to estimate the home potential risk.

Although we could only search for devices exposed on the Internet using the Shodan tool, it would have been useful to have these devices physically for a more detailed study.

A properties and potential risk of the analysed devices

On Table 3 it is shown the properties of more than 60 analysed IoT devices together with the risk value estimated by the model presented in Section 3.

Table 3: Properties and potential risk from the analysed devices

Device	Auth	Auth type	HTTP HTTPS	Default credentials	Exposed	CVE	News	Doc	Type	Score
Phillips HUE	✓	?	HTTP	✖	✖	✓	✓	✓	Lighting system	40
LIFX	?	?	?	?	?	✖	✓	✓	Lighting system	10
GE Link LED	?	?	?	?	?	✖	✓	✓	Lighting system	10
MySmartBlinds	✓	?	?	?	?	✖	✓	✖	Blinds	5
iBlinds	?	?	?	?	?	✖	✓	✖	Blinds	5
Loxane Blinds	✓	user/pass	HTTP	✓	✓	✖	✓	✓	Blinds	57
Nest	✓	Cloud	HTTPS	?	?	✖	✓	✖	Temperature system	10
Momit	✓	Cloud	HTTPS	?	✓	✖	✓	✖	Temperature system	25
Honeywell Centraline	?	?	?	✓	✓	✖	✓	✓	Temperature system	35
Tado	✓	Cloud	HTTPS	?	✖	✖	✖	✓	Temperature system	10
iSmartAlarm	✓	Double Factor	?	?	✖	✖	✖	✓	Alarm system	5
W100 WiFi/PSTN Alarm System	✓	user/pass	?	✓	✖	✖	✖	✓	Alarm system	25
SimpliSafe	✓	user/pass	?	✓	✖	✖	✓	✓	Alarm system	30
Skylink Alarm/Alert System	✓	user/pass	?	?	✖	✖	✖	✖	Alarm system	10
Schlage Sense DeadBolt	✓	?	?	?	✖	✖	✓	✖	Lock	5
August Smart Lock	✓	user/pass	?	✖	✖	✖	✓	✖	Lock	15
Kevo Plus	✓	Double Factor	?	?	✖	✖	✓	✖	Lock	5
Leeo Smart Alert	✓	user/pass	?	✓	✖	✖	✖	✖	Alarm system	20
Samsung LED Smart TV	✓	user/pass	?	?	✓	✓	✓	✓	TV	53
Sony Bravia TV KDL	✓	user/pass	?	?	✓	✓	✓	✓	TV	53
LG SmartTV	✓	user/pass/Pin	?	?	✓	✖	✓	✖	TV	20
EcoATM	✓	user/pass	HTTPS	✖	✓	✖	✖	✖	Paper bin	25
Smart Toy Bear	✓	user/pass	HTTP	✖	✓	✓	✓	✖	Toy	60
Siime Eye	✓	user/pass	HTTP	✓	✖	✖	✓	✖	Toy	37
JAWBONE	✓	Cloud	HTTPS	✖	✖	✖	✓	✖	Wearable	10
SQ-Webcam	✓	user/pass	HTTP	✓	✓	✓	✓	✓	Webcam	75
Webcamxp	✖	user/pass	HTTP	✓	✓	✓	✓	✖	Webcam	95
YawCam	✓	user/pass	HTTP	✓	✓	✓	✓	✖	Webcam	70
Samsung Speaker	✖	?	Bluetooth	✖	✓	✖	✓	✓	Speaker	50
Sonos	✓	user/pass	HTTP	✖	✓	✖	✓	✖	Speaker	42
LG-LAS751M	✓	user/pass	HTTP	✖	✓	✖	✓	✖	Speaker	42
iRobot Roomba 966	✖	?	HTTP	✖	✖	✖	✓	✓	Cleaning robot	47
Robot Dyson	✖	?	HTTP	✖	✖	✖	✓	✓	Cleaning robot	47
LG Hombot	✖	?	HTTP	✖	✖	✖	✓	✓	Cleaning robot	47
Neato Botvac	✓	user/pass	HTTP	✖	✖	✖	✓	✓	Cleaning robot	32
Samsung RF28HMELBSR	✓	App	HTTPS	✖	✖	✖	✓	✖	Fridge	5
Samsung WW10H9600EW	✓	App	HTTPS	✖	✖	✖	✖	✖	Washing machine	0
Bosch Dishwasher with Home Connect	✖	App	HTTPS	✖	✖	✖	✖	✖	Dishwasher	25
Miele Washer-disinfector PG8528	?	?	?	?	?	✓	✓	✓	Dishwasher	28
Samsung Single Wall Oven with Flex Duo	✓	App	HTTPS	✖	✖	✖	✖	✖	Microwave/oven	0
Hornos AGA	✖	App	GSM	✖	✖	✖	✓	✖	Microwave/oven	30
Coffeemakers JURA	✓	App	HTTPS	✖	✖	✓	✖	✖	Coffeemaker	18
Nest Protect	✓	App	HTTPS	✖	✖	✖	✓	✓	Smoke detector	10
Roost Smart Smoke Alarm	✓	App	HTTPS	✖	✖	✖	✖	✖	Smoke detector	0
GreenIQ	✓	Cloud	HTTP	✖	✖	✖	✖	✖	Garden control	17
Z-Wave	✓	user/pass	HTTP	✖	✓	✖	✖	✖	Garden control	37
Netamo	✓	Cloud	HTTPS	✖	✖	✖	✖	✖	Garden control	5
Koubachi	✓	Cloud	HTTP	✖	✖	✖	✖	✖	Garden control	17
Thyssenkrupp	✓	Cloud	HTTPS	✖	✖	✖	✖	✖	Lift	5
KONE elevator	✓	Cloud	HTTPS	✖	✖	✖	✖	✖	Lift	5
Huawei Schindler	✓	Cloud	HTTPS	✖	✖	✖	✖	✖	Lift	5
Jeep	✓	user/pass	HTTPS	✖	✓	✓	✓	✖	Car	48
Aladdin Connect	✓	Cloud	HTTPS	✖	✖	✖	✓	✓	Garage door	15
GarageDoorBuddy	✖	?	HTTP	✓	✖	✓	✖	✖	Garage door	65
Asante	✓	Cloud	HTTPS	✖	✖	✖	✓	✓	Garage door	15
Seagate Central	✖	?	FTP/TELNET	✓	✓	✓	✓	✓	Hard disk	78
Toshiba Canvio	✓	user/pass	HTTP	✓	✓	✖	✖	✓	Hard disk	52
PapercutMF	✓	user/pass	HTTP	✖	✓	✓	✓	✓	Printer	65
Toshiba TopAccess	✖	?	HTTP	✓	✓	✓	✓	✓	Printer	90
Fibaro System	✓	user/pass	HTTP	✓	✓	✖	✖	✓	Home Control	52
Home Wizard	✓	user/pass	App	✓	✓	✖	✖	✓	Home Control	40
Loxone Smart Home	✓	user/pass	HTTP	✓	✓	✖	✖	✓	Home Control	52

Notes

[1] Devices that can be accessed online due to the lack of defensive techniques.
[2] Note that the text in the figures is in Spanish.

References

[1] 2016. IoT worm can hack Philips HUE lightbulbs, spread across cities. https://www.theregister.co.uk/2016/11/10/iot worm can hack philips hue lightbulbs spread across cities/. (2016). [Online; accessed: 15-July-2017].
[2] Architecture of the MEAN Stack 2015. Architecture of the MEAN Stack. Exploring the New Web Dev Tools: MongoDB, ExpressJS, AngularJS, NodeJS. http://shop.oreilly.com/product/0636920039495.do. (2015). [Online; accessed: 15-July-2017].
[3] CVE-2003-1479 2003. CVE Vulnerability CVE-2003-1479. https://cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-2003-1479. (2003). [Online; accessed: 15-July- 2017].
[4] CVE-2004-2094 2004. CVE Vulnerability CVE-2004-2094. https://cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-2004-2094. (2004). [Online; accessed: 15-July- 2017].
[5] CVE-2005-1189 2005. CVE Vulnerability CVE-2005-1189. https://cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-2005-1189. (2005). [Online; accessed: 15-July- 2017].
[6] CVE-2005-1190 2005. CVE Vulnerability CVE-2005-1190. https://cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-2005-1190. (2005). [Online; accessed: 15-July- 2017].
[7] CVE-2008-5674 2008. CVE Vulnerability CVE-2008-5674. https://cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-2008-5674. (2008). [Online; accessed: 15-July- 2017].
[8] CVE-2008-5862 2008. CVE Vulnerability CVE-2008-5862. https://cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-2008-5862. (2008). [Online; accessed: 15-July- 2017].
[9] CVE-2012-1239 2012. CVE Vulnerability CVE-2012-1239. https://cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-2012-1239. (2012). [Online; accessed: 15-July- 2017].
[10] CVE-2012-2210 2012. CVE Vulnerability CVE-2012-2210. https://cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-2012-2210. (2012). [Online; accessed: 15-July- 2017].
[11] CVE-2012-4330 2012. CVE Vulnerability CVE-2012-4330. https://cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-2012-4330. (2012). [Online; accessed: 15-July- 2017].
[12] CVE-2014-1990 2014. CVE Vulnerability CVE-2014-1990. https://cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-2014-1990. (2014). [Online; accessed: 15-July- 2017].
[13] CVE-2014-2657 2014. CVE Vulnerability CVE-2014-2657. https://cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-2014-2657. (2014). [Online; accessed: 15-July- 2017].
[14] CVE-2014-2658 2014. CVE Vulnerability CVE-2014-2658. https://cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-2014-2658. (2014). [Online; accessed: 15-July- 2017].
[15] CVE-2014-2659 2014. CVE Vulnerability CVE-2014-2659. https://cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-2014-2659. (2014). [Online; accessed: 15-July- 2017].
[16] CVE-2014-4883 2014. CVE Vulnerability CVE-2014-4883. https://cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-2014-4883. (2014). [Online; accessed: 15-July- 2017].
[17] CVE-2014-9266 2014. CVE Vulnerability CVE-2014-9266. https://cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-2014-9266. (2014). [Online; accessed: 15-July- 2017].
[18] CVE-2015-5611 2015. CVE Vulnerability CVE-2015-5611. https://cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-2015-5611. (2015). [Online; accessed: 15-July- 2017].
[19] CVE-2015-8039 2015. CVE Vulnerability CVE-2015-8039. https://cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-2015-8039. (2015). [Online; accessed: 15-July- 2017].
[20] CVE-2015-8040 2015. CVE Vulnerability CVE-2015-8040. https://cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-2015-8040. (2015). [Online; accessed: 15-July- 2017].

[21] CVE-2017-7240 2017. CVE Vulnerability CVE-2017-7240. https://cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-2017-7240. (2017). [Online; accessed: 15-July- 2017].

[22] Fridgevuln 2015. Smart refrigerator hack exposes Gmail login credentials. http://www.networkworld.com/article/2976270/internet-of-things/ smart-refrigerator-hack-exposes-gmail-login-credentials.html. (2015). [On- line; accessed: 15-July-2017].

[23] Gartner. 2013. Gartner Says the Internet of Things Installed Base Will Grow to 26 Billion Units By 2020. http://www.gartner.com/newsroom/id/2636073. (2013). [Online; accessed: 15-July-2017].

[24] HUE 2016. Philips HUE - Personal Wireless Lighting. http://www2.meethue. com/en-gb/. (2016). [Online; accessed: 15-July-2017].

[25] ICS 2015. ICS - Professional services. http://www.ics.com/sites/default/ files/ images/ioe.png. (2015). [Online; accessed: 15-July-2017].

[26] ITU. 2013. Internet of Things Global Standards Initiative. http://www.itu.int/en/ ITU-T/gsi/iot/Pages/default.aspx. (2013). [Online; accessed: 15-July-2017].

[27] C. Liu, Y. Zhang, J. Zeng, L. Peng, and R. Chen. 2012. Research on Dynamical Security Risk Assessment for the Internet of Things inspired by immunology. (2012). DOI: http://dx.doi.org/10.1109/ICNC.2012.6234533

[28] Loxone Blind 2017. Loxone Blind Control. https://www.loxone.com/enen/ smart-home/blinds/. (2017). [Online; accessed: 15-July-2017].

[29] Shafqat Mehmood. 2016. Enterprise Survival Guide for Ransomware Attacks. (2016).

[30] Yair Meidan, Michael Bohadana, Asaf Shabtai, Martin Ochoa, Nils Ole Tippen- hauer, Juan Davis Guarnizo, and Yuval Elovici. 2017. Detection of Unauthorized IoT Devices Using Machine Learning Techniques. (2017). arXiv:arXiv:1709.04647

[31] Diego M. Mendez, Ioannis Papapanagiotou, and Baijian Yang. 2017. Internet of Things: Survey on Security and Privacy. (2017). http://arxiv.org/abs/1707.01879 [32] Miele 2016. Washer-desinfector Miele. https://www.miele.co.uk/professional/ large-capacity-washer-disinfectors-560.htm?mat=10339600&name=PG 8528#item-2-2. (2016). [Online; accessed: 15-July-2017].

[32] Mujahid Mohsin, Muhammad Sardar, Osman Hasan, and Zahid Anwar. 2017.

[33] IoTRiskAnalyzer : A Probabilistic Model Checking Based Framework for Formal Risk Analytics of the Internet of things. (2017).

[34] Papercutmf 2016. PaperCutMF. https://www.papercut.com/products/mf/. (2016). [Online; accessed: 15-July-2017].

[35] Pierluigi Paganini. 2014. Philips SmartTV susceptible to serious hack ac cording ReVuln experts. http://securityaffairs.co/wordpress/23523/hacking/ philips- smarttv- susceptible-serious- hack- according- revuln- experts.html. (2014). [Online; accessed: 15-July-2017].

[36] Pierluigi Paganini. 2017. The software engineer Darren Cauthon reported his LG Smart TV was infected with ransomware on Christmas day, the malware asked for $500 to unlock the device. http://securityaffairs.co/wordpress/54991/malware/ lg-smart-tv-ransomware.html. (2017). [Online; accessed: 15-July-2017].

[37] Shodan 2009. The Search Engine for Internet connected devices. https://www. shodan.io. (2009). [Online; accessed: 15-July-2017].

[38] Siime 2017. This New Dildo Camera Can Be Hacked Way Too Easily. http://uk. complex.com/life/2017/04/dildo-camera-can-be-hacked-way-too-easily. (2017). [Online; accessed: 15-July-2017].

[39] SiimeSold 2016. Siime Eye. https://www.touchofmodern.com/sales/ svakom-cfe31d24-d840-4095-a4a5-d148b783260a/siime-eye-violet. (2016). [On- line; accessed: 15-July-2017].

[40] Sqwebcam 2015. SQ-WebCam. https://sourceforge.net/projects/sqcam/. (2015). [Online; accessed: 15-July-2017].

[41] ThosibaPrinter 2016. Toshiba TopAccess Printer. http://www.toshibatec.am/en/software/58. (2016). [Online; accessed: 15-July-2017].

[42] Toybears 2016. Smart Toybear. http://smarttoy.com/products. (2016). [Online; accessed: 15-July-2017].

[43] toyhack 2017. Internet of Things Teddy Bear Leaked 2 Million Parent and Kids Message Recordings. https://goo.gl/3dDZrV. (2017). [Online; accessed: 15-July-2017].

[44] TVHacking 2017. Here's How the CIA Allegedly Hacked Samsung Smart TVs – And How To Protect Yourself. https://www.forbes.com/sites/thomasbrewster/2017/03/07/cia-wikileaks-samsung-smart-tv-hack-security. (2017). [Online; accessed: 15-July-2017].

[45] vulndishwasher 2017. Dishwasher has directory traversal bug. https: //www.theregister.co.uk/2017/03/26/miele joins internetofst hall of shame/ (2017). [Online; accessed: 15-July-2017].

[46] Webcamxp 2016. WebCamXP. http://www.webcamxp.com/home.aspx. (2016). [Online; accessed: 15-July-2017].

[47] Wired 2015. Hackers Remotely kill a Jeep on the Highway. https://www.wired.com/2015/07/hackers-remotely-kill-jeep-highway/. (2015). [Online; accessed: 15-July-2017].

[48] Yawcam 2013. YawCam. http://www.yawcam.com. (2013). [Online; accessed: 15-July-2017].

Cycle Structure and Reachability Analysis for Cipher Spritz with Small *N*

JÖRG KELLER

Abstract Spritz has been proposed as a replacement for RC4. For embedded applications that use it as a stream cipher or pseudo-random number generator with smaller parameter N than the standard $N = 256$, the choice of N should be as small as possible for performance, but large enough to provide sufficient security. Hence, we investigate which fraction of the state space is reachable with keysetup and subsequent output, and which cycle length can be expected for small N . Next to some elementary theoretical investigations, we experimentally do an exhaustive search on the state space for $N = 4, 6, 8$.

Keywords: • Spritz Cipher • State Space Analysis • Pseudo-Random Number Generator • Security Analysis • Secure Embedded Systems •

CORRESPONDENCE ADDRESS: Jörg Keller, Ph.D., Professor, FernUniversität in Hagen, Faculty of Mathematics and Computer Science, Universitätsstr. 1, 58084 Hagen, Germany, e-mail: joerg.keller@fernuni-hagen.de.

DOI https://doi.org/10.18690/978-961-286-114-8.12 ISBN 978-961-286-114-8

Available at: http://press.um.si.

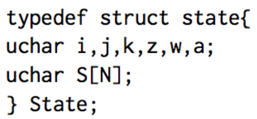
1 Introduction

RC4 has been a very popular cipher but should not be used anymore do to weaknesses [5]. In 2014, Rivest and Schuldt proposed Spritz as a replacement for RC4 which works along similar lines but avoids RC4's weaknesses [6]. Several authors have investigated the security of Spritz [1, 2], but to our knowledge, there have been no investigatons about the fraction of the state space that is reachable with a key setup when the cipher is used as a stream cipher or pseudo-random number generator (PRNG). For the standard parametrization with $N = 256$ the state space size of $n = N^6 \cdot N!$ is so huge that sufficient security can be assumed even if only a fraction of the state space is accessible. In embedded systems however, a much smaller parameter N might be chosen for reasons performance. For example, N = 16 would allow an implementation where each state variable only comprises a nibble, i.e. 4 bit. For security reasons, the exact choice of N then starts to depend on the fraction of states that can be reached, and on the period lengths that can be relied on, e.g. when using Spritz as a stream cipher or pseudo-random number generator. We investigate both questions by first doing some elementary theoretical considerations (Sect. 2) and then doing an exhaustive exploration of the state space for $N = 4, 6, 8$ (Sect. 3), focusing on cycle lengths and which cycles are reachable from a start state via a key setup. In Sect. 4 we give conclusions and an outlook to future work.

2 Basics

2.1 Spritz Cipher

The Spritz cipher [6] is a drop-in replacement for the widely-used cipher RC4. It strives to handle key setup differently, and has more variables in its state, to overcome known weaknesses of RC4. Spritz is formulated as a Sponge function [3]. Thus, among others, it can be used as a pseudo-random number generator. While key setup has been changed, generation of output streams still happens quite similarly to RC4.



```
typedef struct state{
uchar i,j,k,z,w,a;
uchar S[N];
} State;
```

Figure 1: State declaration for Spritz

Spritz is really a family of functions parameterized in an integer , where the default value is $N = 256$. e internal state consists of six integer variables in the range $\{0, \ldots, N - 1\}$, which are named i, j, k, z, w, a, and of a permutation $S \in S_N$, where S_N is the set of permutations on N elements. We denote the size of the state space as s_N. Figure 1 shows the state as a struct declaration. As N normally is 256 at most, the variables are declared as unsigned characters. Please note that strictly speaking, also the chosen value of N should be part of the state. However, as typical implementations are optimized for one particular , this is skipped.

The initialization (see function InitializeState) starts with $i = j = k = z = a = 0$, $w = 1$ and $S = \text{id}$. Then, a key of arbitrary length, consisting of symbols in the range $\{0, \ldots, N -$

$1\}$, is read in, and the state is transformed by each symbol in function $\mathrm{AbsorbByte}$. This function calls a sequence of other functions ($\mathrm{Shuffle}, \mathrm{Whip}, \mathrm{Crush}, \mathrm{Update}$), which modify a, *w*, i, *j*, *k* and *S*, with the constraints that *w* and *N* must be always co-prime, and that a must be zero at the end of the key setup. If it is not, there is one call to $\mathrm{Shuffle}$ at the start of output (cf. Alg. 2), which sets $a = 0$. The functions for key setup are shown in Alg. 1. Please note that in all functions, the state parameter is passed as call by reference, i.e. in an implementation it would really be a pointer, and state modifications within a function are passed back to the calling function.

After initialization, when used as a pseudo-random number generator, Spritz generates symbols by calls to Drip which updates the state in function Update by modifying variables i, *j*, *k* and *S*, and outputs a symbol in the range $\{0, \ldots, N - 1\}$ in function $Output$, which modifies variable *z*. The transition from one state to the follow-up state is bijective [6]. The functions for output generation are shown in Alg. 2. For a complete listing and discussion of all functions, we refer the reader to [6].

Algorithm 1 Spritz key setup functions.

Precondition: st is current state, call by reference

```
function InitializeState($st, N$)
    $st.i, st.j, st.k, st.z, st.a \leftarrow 0$
    $st.w \leftarrow 1$
    for $v \leftarrow 0$ to $N - 1$ do
        $st.S[v] \leftarrow$ v

function Absorb($st, I$)
    for $v \leftarrow 0$ to $|I| - 1$ do
        AbsorbByte($st, I[v]$)

function AbsorbByte($st, b$)
    AbsorbNibble($st$,low(b))    ▷ low, high do mod and div $\sqrt{N}$
    AbsorbNibble($st$,high(b))

function AbsorbNibble($st, x$)
    if $st.a = \lfloor N/2 \rfloor + x\rfloor$ then
        Shuffle($st$)
    Swap($st.S[st.a], st.S[st.a = \lfloor N/2 \rfloor + x]$)
    $st.a \leftarrow st.a + 1$

function Shuffle($st$)
    Whip($st, 2N$)
    Crush($st$)
    Whip($st, 2N$)
    Crush($st$)
    Whip($st, 2N$)
    $st.a \leftarrow 0$

function Whip($st, r$)
    for $v \leftarrow 0$ to $r - 1$ do
        Update($st$)    ▷ Part of output functions, cf. Alg. 2
    repeat
        $st.w \leftarrow st.w + 1$
    until gcd$(st.w, N) = 1$

function Crush($st$)
    for $v \leftarrow 0$ to $\lfloor N/2 \rfloor - 1$ do
        if $st.S[v] > st.S[N - 1 - v]$ then
            Swap($st.S[v], st.S[N - 1 - v]$)
```

2.2 Graphs of Random Bijective Functions

If we have a PRNG with state space M, and state transition function $f : M \rightarrow M$, then this induces a directed graph $G = (V, E)$ with $V = M$ and $E = \{(v, f(v)) | \upsilon \in \mathrm{M}\} \subset M \times M$. Each node has exactly one outgoing edge which leads to the node of the follow-up state. In the case of Spritz, state transition is bijective, and hence the graph consists only of cycles.

As PRNGs should not expose a regular structure in their state transition, the state transition graph should resemble a graph for a randomly chosen function. As Spritz is bijective, we must compare with a randomly chosen bijective function, i.e. a random permutation. If we choose a permutation randomly and equidistributed among all possible permutations on s_N elements, then the expected length of the longest cycle is $(1 - 1/e)\ s_N \approx 0.63 \cdot s_N$, the number of cycles is expected to be around $ln(s_N)$, and cycle lengths on average shrink in a ratio of 2 : 1 [7]. The structure of a concrete PRNG should not deviate too much from this.

Algorithm 2 Spritz output functions.

Precondition: st is current state, call by reference

```
function DRIP($st$)
    if $st.a > 0$ then                ▷ At most once after setup
        SHUFFLE($st$)
    UPDATE($st$)
    return UPDATE($st$)

function SETUP($st$)
    $st.i \leftarrow st.i + st.w$
    $st.j \leftarrow st.k + st.S[st.j + st.S[st.i]]$
    $st.k \leftarrow st.i + st.k + st.S[st.j]$
    SWAP($st.S[st.i], st.S[st.j]$)

function OUTPUT($st$)
    $st.z \leftarrow st.S[st.j + st.S[st.i + st.S[st.z + st.k]]]$
    return $st.z$
```

3 Analysis of spritz

3.1 Theoretical Considerations

Rivest and Schuldt state that “Spritz has at most $\#(N) = N^6 N!$ states.” [6, p. 7]. We note that after keysetup, i.e. after InitializeState, Absorb and a possible call to Shuffle in the first invocation of Drip, variable $a = 0$ either as a result of Absorb or the extra Shuffle. Also, we note that variable w always has a value such that $\gcd(w, N) = 1$, as it is initialized to $w = 1$ and only modified in Whip with that condition. As during the calls to Drip, i.e. in Update and Output, only the permutation S and variables i, j, k, and z are modified, the number of states to be used during output is bounded by

$$s_N = N^4 \cdot \varphi(N) \cdot N!\,,$$

where $\varphi(N)$ denotes the Euler function that states how many integers in $\{1, \ldots, N - 1\}$ are co-prime to N.

When we investigate the cycle structure of Spritz during output of pseudo-random numbers, we must compare the resulting cycle lengths with expected values for cycles lengths in graphs of random bijective functions on s_N elements. However, as w is set at one of the $\varphi(N)$ possible values during key setup, and not modified during output, the graph partitions into $\varphi(N)$ isomorphic graphs of $s_N/\varphi(N)$ elements each, so that we can expect somewhat smaller cycle lengths.

Furthermore, we note that while i, j, k and S are modified during key setup (Whip calls Update which modifies i, j, k and S, and S is also modified in Absorbnibble and Crush) $z = 0$ during keysetup, as z is only modified in output. Thus, the states that can be reached as the result of a key setup (we will call them entry states or ES) can be characterized by $a = z = 0$ and $\gcd(w, N) = 1$, and their number is bound by

$$e_N = N^3 \cdot \varphi(N) \cdot N!\,,$$

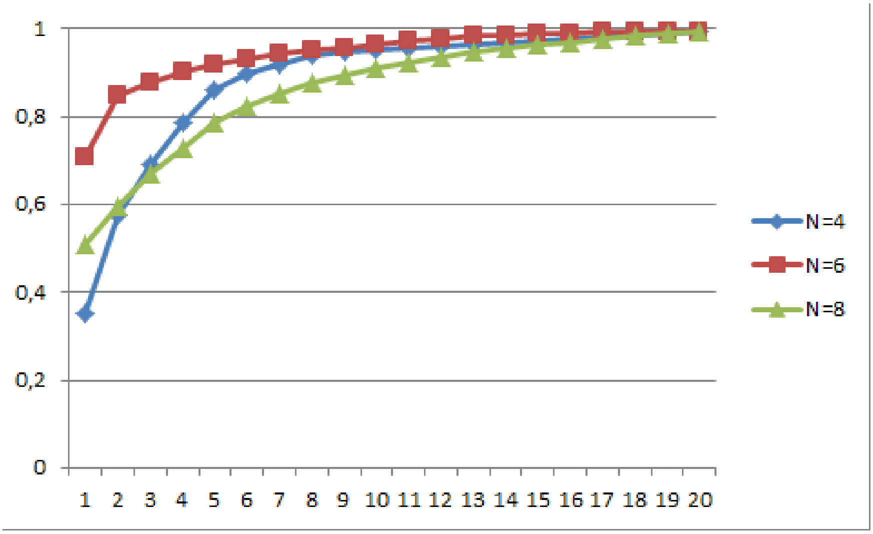


Figure 2: Aggregated cycle lengths of longest cycles for each *N*, as fraction of $s_n/\varphi(N)$.

When we consider to use keys of length at most k, then there are at most $\sum_{i\leq k} N^i = (N^{k+1}-1)/(N-1)$ keys and thus ES reachable. us, to have a chance to reach all e_N entry states, k must be on the order of N. Please note that if we assume that the entry states are more or less evenly distributed over the cycles, a much smaller number of entry states suffices to reach at least the larger cycles, and thus a notable fraction of all states. However, if the number of entry states is small, then a cycle, even if it is long, can be entered only at a small number of entry states, which might be distinguishable by their subsequent output patterns, so that internal states can be derived.

3.2 Cycle Structure

We compute the lenghts of all cycles that contain at least one entry state, i.e. that can in principle be reached by a key setup. For $N = 8$, there are 16 states that are not on a cycle with an entry state. We only consider the cycle structure for $w = 1$, as the cycle lengths for other values of w are identical. ere 24, 60, and 120 cycles for $N = 4, 6, 8$, respectively, which is notably higher than $\ln(s_N/\varphi(N))$. The longest cycles have lengths 2172, 660390 and 84143080, resp., which is about 0.3, 0.65, 0.5 when compared to $s_N/\varphi(N)$, i.e. shorter than expected but still su ciently long. e aggregated sizes of the 20 largest cycles (as fraction of $s_n/\varphi(N)$) for each N are depicted in Fig. 2.

3.3 Reachability

We have checked for each N , how many cycles and states can be reached via a key setup with all keys of length up to k . For example, with $N = 4$, there are 4 different keys of length 1, 16 keys of length 2, and 64 keys of length 3 (of which 2 keys already lead to the same entry state). These keys target partly the same, and partly different cycles. Fig. 3 depicts the number of states on cycles reachable with key setups of lengths up to 10, for the different N , as fraction of s_N (as the key setups may lead to any value of w co-prime to N , we had to consider the complete state space.) It is clearly visible that already for short key lengths with $k \geq 4$, for all N a fraction of more than 90% of the elements can be reached (comprised from the largest cycles), but that longer key lengths do not bring an advantage in reachability. The difference is still, which entry states on a long cycle can be reached, i.e. diversity of the subsequent output sequences.

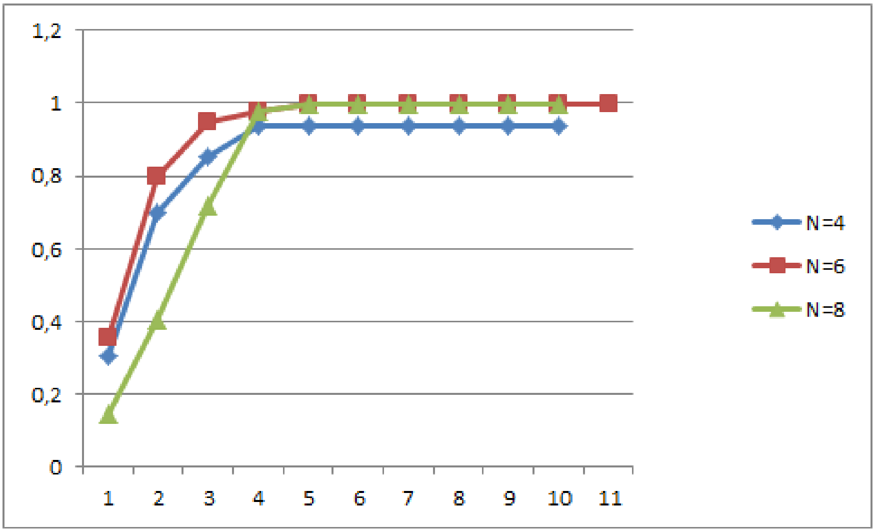


Figure 3: States reachable with keys of length up to 10, for each *N*, as fraction of s_n.

3.4 Notes on Implementation

The implementation was done as a sequential program in the programming language C. The main data structure is a bit vector which marks entry states as visited. For the cycle structure analysis, we enumerated all entry states for a particular *w*, and for each unvisited entry state (i.e. a newly detected cycle) we followed the cycle until we reached this entry state again, marking all states on the cycle as visited. The algorithm, which is a depth first search tailored to the specific graph structure, is given in Alg.

Algorithm 3 Cycle structure analysis

```
function CYCLEANALYZE($w$)
    for all entry states $es$ with given $w$ do
        mark $es$ as unvisited
    for all entry states $es$ with given $w$ do
        if $es$ is unvisited then
            $st \leftarrow es$
            $M \leftarrow \emptyset$                    ▻ Set of entry states per cycle
            $l \leftarrow 0$                            ▻ Counter for cycle length
            while $st$ is unvisited do
                $M \leftarrow M \cup \{st\}$
                mark $st$ as visited
                repeat
                    DRIP($st$)
                    $l \leftarrow l + 1$
                until $st$ is entry state
            Save "Cycle of length $l$with entry states in $M$"
```

For the reachability, we used this information and enumerated all possible keys of increasing length, performed the key setup for each key, and checked which entry states (and thus which cycles) are reachable.

4 Conclusions

We have experimentally demonstrated that the cycle structure of Spritz with small *N* to some extent resembles the structure of a random permutation, at least for each fixed value of parameter *w*. We have further demonstrated that a large fraction of the states (more than 90% in all cases) can be reached via key setup with reasonably long keys. Hence, Spritz can be used in embedded systems with moderate security requirements. Our future work will comprise to explore Spritz for *N* = 10 to 16, which will necessitate a massive parallelization in the spirit of [4], and for which an implementation on a graphics processing unit (GPU) is underway.

Acknowledgments

The author would like to thank Mr. Stefan Lange for discussions about implementing Spritz.

References

[1] Ralph Ankele, Stefan Ko ̈lbl, and Christian Rechberger. 2015. State-Recovery Analysis of Spritz. In Progress in Cryptology – LATINCRYPT 2015: 4th International Conference on Cryptology and Information Security in Latin America, Guadalajara, Mexico, August 23-

26, 2015, Proceedings. Springer International Publishing, Cham, 204–221. DOI: http://dx.doi.org/10.1007/978-3-319-22174-812

[2] Subhadeep Banik and Takanori Isobe. 2016. Cryptanalysis of the Full Spritz Stream Cipher. In Fast Software Encryption: 23rd International Conference, FSE 2016, Bochum, Germany, March 20-23, 2016, Revised Selected Papers. Springer Berlin Heidelberg, Berlin, Heidelberg, 63–77. DOI: http://dx.doi.org/10.1007/ 978-3-662-52993-54

[3] Guido Bertoni, Joan Daemen, Michael Peeters, and Gilles Van Assche. 2008. On the indifferentiability of the sponge construction. In Proc. Eurocrypt, LNCS 4965. Springer, 181–197.

[4] Jorg Keller and Jop F. Sibeyn. 2001. Beyond External Computing: Analysis of the Cycle Structure of Permutations. In Euro-Par 2001: Parallel Processing, 7th International Euro-Par Conference Manchester, UK August 28-31, 2001, Proceedings. 333–342. DOI: http://dx.doi.org/10.1007/3-540-44681-8 48

[5] Andreas Klein. 2008. A acks on the RC4 stream cipher. Designs, Codes and Cryptography 48, 3 (01 Sep 2008), 269–286. DOI: http://dx.doi.org/10.1007/ s10623- 008- 9206- 6

[6] Ronald L. Rivest and Jacob C. N. Schuldt. 2014. Spritz—A spongy RC4-like stream cipher and hash function. In Presented at Charles River Crypto Day (2014-10-24).

[7] Robert Sedgewick and Philippe Flajolet. 1996. Introduction to the Analysis of Algorithms. Addison-Wesley.