Network Architectures
and Services
NET 2015-03-1

Proceedings

Proceedings of the Seminars Future Internet (FI) and Innovative Internet Technologies and Mobile Communications (IITM)

Winter Semester 2014/2015

29. 10. 2014 – 22. 02. 2015

Munich, Germany

Editors

Georg Carle, Daniel Raumer, Lukas Schwaighofer

Publisher

Chair for Network Architectures and Services

Department of Informatics

Technische Universität München

Network Architectures
and Services
NET 2015-03-1

FI & IITM
WS 14/15

Proceedings zu den Seminaren Future Internet (FI) und Innovative Internet Technologien und Mobilkommunikation (IITM)

Wintersemester 2014/2015

München, 29. 10. 2014 – 22. 02. 2015

Editoren: Georg Carle, Daniel Raumer, Lukas Schwaighofer

Organisiert durch den Lehrstuhl Netzarchitekturen und Netzdienste (I8),
Fakultät für Informatik, Technische Universität München

Technische Universität München

Proceedings of the Seminars
Future Internet (FI), and Innovative Internet Technologies and Mobile Communication Networks (IITM)
Winter Semester 2014/2015

Editors:

Georg Carle
Lehrstuhl Netzarchitekturen und Netzdienste (I8)
Technische Universität München
D-85748 Garching b. München, Germany
E-mail: carle@net.in.tum.de
Internet: http://www.net.in.tum.de/~carle/

Daniel Raumer
Lehrstuhl Netzarchitekturen und Netzdienste (I8)
E-mail: raumer@net.in.tum.de
Internet: http://www.net.in.tum.de/~raumer/

Lukas Schwaighofer
Lehrstuhl Netzarchitekturen und Netzdienste (I8)
E-mail: schwaighofer@net.in.tum.de
Internet: http://www.net.in.tum.de/~schwaighofer/

Cataloging-in-Publication Data

Seminars FI & IITM WS 14/15
Proceedings zu den Seminaren „Future Internet“ (FI) und „Innovative Internettechnologien und Mobilkkommunikation“ (IITM)
München, Germany, 29. 10. 2014 - 22. 02. 2015
ISBN: 978-3-937201-47-4

ISSN: 1868-2634 (print)
ISSN: 1868-2642 (electronic)
DOI: 10.2313/NET-2015-03-1
Lehrstuhl Netzarchitekturen und Netzdienste (I8) NET 2015-03-1
Series Editor: Georg Carle, Technische Universität München, Germany

Vorwort

Vor Ihnen liegen die Proceedings des Seminares „Future Internet“ (FI) und des Seminares „Innovative Internettechnologien und Mobilkommunikation“ (IITM). Wir sind stolz Ihnen Ausarbeitungen zu interessanten Themen, die im Rahmen unserer Seminare im Wintersemester 2014/2015 an der Fakultät für Informatik der Technischen Universität München verfasst wurden, präsentieren zu dürfen. Den Teilnehmerinnen und Teilnehmern stand es, wie in der Vergangenheit auch schon, frei das Paper und den Vortrag in englischer oder in deutscher Sprache zu verfassen. Dementsprechend finden sich sowohl englische als auch deutsche Paper in diesen Proceedings.

Unter allen Themen, die sich mit Aspekten der Computernetze von morgen befassen, verliehen wir in jedem der beiden Seminare einen Best Paper Award. Im FI Seminar ging dieser an Herrn Lukas Märdian, der in seiner Ausarbeitung „What's New in the Linux Network Stack?“ die Entwicklung der Netzwerkfunktionalität des Linux Kernels anschaulich beschreibt und bewertet. Im IITM Seminar erhielt Herr Felix Emmert den Award. Die von ihm verfassten Ausarbeitung „Out-of-Band Network Management“ beschreibt den Stand der Technik des Servermanagements aus der Ferne und beleuchtet Sicherheitsaspekte auf anschauliche Weise.

Einige der Vorträge wurden aufgezeichnet und sind auf unserem Medienportal unter `http://media.net.in.tum.de` abrufbar.

Im Seminar FI wurden Beiträge zu aktuellen Themen der Netzwerkforschung vorgestellt. Die folgenden Themen wurden abgedeckt:

- Attack Taxonomies and Ontologies
- Freenet
- Timing of Cyber Conflict
- Exploring DDoS Defense Mechanisms
- What's New in the Linux Network Stack?
- Internet science-Creating better browser warnings
- Hardware-accelerated Galois Field Arithmetic on the ARMv8 Architecture
- Measuring Privacy
- Internet Science – Critical Infrastructures

Auf `http://media.net.in.tum.de/#%23Future%20Internet%23WS14` können die aufgezeichneten Vorträge zu diesem Seminar abgerufen werden.

Im Seminar IITM wurden Vorträge aus dem Themenbereich der Netzwerktechnologien inklusive Mobilkommunikationsnetze vorgestellt. Die folgenden Themen wurden abgedeckt:

- Out-of-Band Network Management
- Smart Buildings vs. Data Privacy Law
- Moving Target Defense
- Cryptocurrency Brings New Battles into the Currency Market

Auf `http://media.net.in.tum.de/#%23IITM%23WS14` können die aufgezeichneten Vorträge zu diesem Seminar abgerufen werden.

Wir hoffen, dass Sie den Beiträgen dieser Seminare wertvolle Anregungen entnehmen können. Falls Sie weiteres Interesse an unseren Arbeiten habe, so finden Sie weitere Informationen auf unserer Homepage `http://www.net.in.tum.de`.

München, März 2015

Georg Carle

Daniel Raumer

Lukas Schwaighofer

Preface

We are pleased to present you the interesting program of our seminars on "Future Internet" (FI) and "Innovative Internet Technologies and Mobil Communication" (IITM) which took place in the winter semester 2014/2015. In both seminar courses the authors were free to write their paper and give their talk in English or German.

In both of the seminars we honored the best paper with an award. In the FI seminar the award was given to Mr Lukas Märdian who discussed recent changes in the Linux networking code in his Paper "What's New in the Linux Network Stack?". In the IITM seminar Mr Felix Emmert was honored with the Best Paper Award. His paper "Out-of-Band Network Management" presented the State-of-the-Art in out-of-band server management and discussed security issues.

Some of the talks were recorded and published on our media portal `http://media.net.in.tum.de`.

In the seminar FI we dealt with issues and innovations in network research. The following topics were covered:

- Attack Taxonomies and Ontologies
- Freenet
- Timing of Cyber Conflict
- Exploring DDoS Defense Mechanisms
- What's New in the Linux Network Stack?
- Internet science-Creating better browser warnings
- Hardware-accelerated Galois Field Arithmetic on the ARMv8 Architecture
- Measuring Privacy
- Internet Science – Critical Infrastructures

Recordings can be accessed on `http://media.net.in.tum.de/#%23Future%20Internet%23WS14`.

In the seminar IITM we dealt with different topics in the area of network technologies, including mobile communication networks. The following topics were covered:

- Out-of-Band Network Management
- Smart Buildings vs. Data Privacy Law
- Moving Target Defense
- Cryptocurrency Brings New Battles into the Currency Market

Recordings can be accessed on `http://media.net.in.tum.de/#%23IITM%23WS14`.

We hope that you appreciate the contributions of these seminars. If you are interested in further information about our work, please visit our homepage `http://www.net.in.tum.de`.

Munich, March 2015

Seminarveranstalter

Lehrstuhlinhaber

Georg Carle, Technische Universität München, Germany (I8)

Seminarleitung

Daniel Raumer, Technische Universität München, Germany
Lukas Schwaighofer, Technische Universität München, Germany

Betreuer

Paul Emmerich (emmericp@net.in.tum.de)
Technische Universität München, Mitarbeiter I8

Oliver Gasser (gasser@net.in.tum.de)
Technische Universität München, Mitarbeiter I8

Stephan Günther (guenther@net.in.tum.de)
Technische Universität München

Nadine Herold (herold@net.in.tum.de)
Technische Universität München, Mitarbeiterin I8

Holger Kinkelin (kinkelin@net.in.tum.de)
Technische Universität München, Mitarbeiter I8

Marcel von Maltitz (maltitz@net.in.tum.de)
Technische Universität München, Mitarbeiter I8

Heiko Niedermayer (niedermayer@net.in.tum.de)
Technische Universität München, Mitarbeiter I8

Daniel Raumer (raumer@net.in.tum.de)
Technische Universität München, Mitarbeiter I8

Johann Schlamp (schlamp@net.in.tum.de)
Technische Universität München, Mitarbeiter I8

Seminarhomepage

`http://www.net.in.tum.de/de/lehre/ws1415/seminare/`

Inhaltsverzeichnis

Seminar Future Internet

Seminar Innovative Internet Technologien und Mobilkommunikation

Attack Taxonomies and Ontologies

Natascha Abrek
Advisor: Nadine Herold
Seminar Future Internet SS2014
Chair for Network Architectures and Services
Department of Informatics, Technical University of Munich
Email: abrek@in.tum.de

ABSTRACT

In the past few years network security threats have increased significantly. Methods for attacks have not only grown in diversity but also became more sophisticated. The increased need for security mechanisms and countermeasures requires a comprehensive understanding of those attacks and their characteristics. To organize the knowledge of attacks a large variety of classifications were proposed in form of taxonomies and ontologies. The development of these classifications has emerged as an effective means for developing awareness systems and creating common descriptive languages. However, due to the high diversity of attacks no standard classification of network attacks exists so far. In this paper, a survey of existing attack taxonomies and ontologies is presented to create an overview of conducted work in this field of research. Furthermore, issues and drawbacks are discussed in a comparative analysis. The conducted survey has shown the need of a flexible, standardized classification of attacks and vulnerabilities to enable efficient knowledge sharing among the scientific community.

Keywords

Attack taxonomy, security ontology

1. INTRODUCTION

Latest publicized cyber-attacks against corporate and public organizations highlight the persistent threat against network security. The variety of methods to target personal, corporate or financial information has significantly increased and attacks became more sophisticated. New network vulnerabilities and attack possibilities were discovered by aggressors. New developments such as blended threats and information warfare techniques evolved. To protect from this wave of network threats robust countermeasures are necessary. However, the development of such security measures requires a comprehensive understanding of network attacks and their classifications. Taxonomies help to classify threats into well-defined categories. Bishop and Bailey [2] define a taxonomy as a system of classification which allows the unique identification of objects. Taxonomies help to organize knowlege and can serve as a helpful tool in the modeling process of system security and security policies. In the past, there have been numerous attempts to develop attack taxonomies [13, 18, 23, 27]. They range from general taxonomies to taxonomies which cover specific application domains or attack fields [12, 6]. Although numerous taxonomies have been introduced in the literature, no standard classification was developed so far. Research has shown several other drawbacks of taxonomies as well. The lack in consistency and extensibility makes them deficient in defining semantic relationships. In addition, the hierarchical order in taxonomies limit the possibilities of reuseability[26]. Therefore, the transition to ontologies is neccessary. Although both concepts are similar, the main difference is that an ontology complements the hierarchical order of a taxonomy with additional relationships. According to Gruber [11] an ontology is an explicit specification of a conceptualization. Ontologies represent powerful means to organize and represent knowledge in a structured and formal way. Additionally, they ease the process of communication and knowledge sharing [26]. Already several ontologies were developed in the area of network security. Still, also in the field of attack ontologies the development of a consistent ontology has not been accomplished so far. In this paper, a systematic survey of existing literature on attack taxonomies and ontologies is conducted. Thereby, two representing taxonomies and ontologies are selected and discussed. The selected papers cover research of network attacks in general. Classifications with aspect to specific fields were not considered. Furthermore, the focus lies on research conducted in recent years, reducing the selection to research papers published between 2012 and 2014. Following, a systematic analysis is carried out comparing the most relevant aspects. Goal of this work is to create an overview of conducted research in this field and to help researchers to take further steps towards a standardized classification of network attacks.

The remainder of this paper is organized as follows. In Section 2 characteristics of good taxonomies are discussed which also build the criteria for the following analysis. Furthermore, two selective taxonomies are presented. Section 3 covers the benefits of the transition from taxonomies to ontologies and the presentation of two existing security ontologies. In Section 4 an analysis is conducted discussing differences, advantages as well as disadvantages of the presented taxonomies and ontologies according to the defined criteria. Section 5 shows an overview of related literature surveying existing attack taxonomies and ontologies. Finally, in Section 6 the conclusion of this survey is presented.

2. ATTACK TAXONOMIES

In the field of network and computer security a great number of taxonomies classifying security threats and vulnerabilities were developed. In the following section, first the main characteristics of sufficient taxonomies are described. Then two selected attack taxonomies are presented in detail.

doi: 10.2313/NET-2015-03-1_01

2.1 Characteristics of a taxonomy

While computer and network attacks have become a consistent threat, the methods used to describe them are often inconsistent. In addition, the attack classification and detection represents a challenging task due to the highly increased number of threats during the years. That is why classification schemes such as taxonomies are pervasive means in the field of computer and network security engineering. The objective of a taxonomy is to provide a consistent instrument to classify attacks based on their characteristics. For an attack to be launched, security vulnerabilities are exploited. A vulnerability is a security exposure which results from flaws in a system or code. By providing an overview of attack characteristics such as vulnerabilities, a taxonomy can serve as a helpful tool in the security modeling process and in security assessment. Attack detection systems like intrustion detection systems can make use of taxonomies to identify a threat by the defined characteristics.

Before examining existing taxonomies, the main characteristics of a taxonomy have to be discussed. A taxonomy organizes classes in a hierarchical manner. The hierarchy is structured in multiple levels representing the depth of classification. The relationships between classes and subclasses are realized with the *is-a* relationship. In fact, this is the only relationship that can be drawn between classes in a taxonomy. Researchers have summerized a number of characteristics a taxonomy should satisfy in order to be sufficient. General requirements towards a taxonomy include the following:

Accepted: The structure of the taxonomy has to be intuitive and logical so that it can be easily understood and generally approved. It should build on previous, well-known research [15].
Comprehensible: The taxonomy should be understandable to both experts as well as those with less expertise. The concept has to be presented in a concise and clear form [19].
Determined: A clear definition and explanation for the developed classification is to be provided [17].
Exhaustive: A taxonomy is considered exhaustive or complete if all possibilities of attacks are accounted for [15].
Mutually exclusive: To achieve a mutual exclusive taxonomy every attack should be categorized into only one category. The developed categories must not overlap [15].
Repeatable: If the taxonomy is applied repeatedly, it has to result in the same classification [15].
Terms well defined: Only established security terminology should be used in the taxonomy. This is neccessary to avoid confusion and to build on previous, general knowledge [19].
Unambiguous: A precise definition of the categories is neccessary to prevent an ambiguous or unclear classification of an attack [15].
Useful: A taxonomy is useful when it is used to gain insights into a specific field of study [15].

According to Hansman [13] it is not possible or even neccessary for a taxonomy to fulfill all requirements at the same time. The degree on which a taxonomy aims to meet the requirements depends on the particular goal of the taxonomy. Authors have also identified a few more characteristics such as objectivity, appropriateness or primitivity [17, 1]. However, these characteristics are not taken into account in this survey since the presented taxonomies address only the above mentioned characteristics. The same characteristics will later on serve as criteria to conduct a comparative analysis between the presented taxonomies. In the following of this section, two selected attack taxonomies are presented. For a better understanding of the classification process with these taxonomies, they will be applied to a selected attack, the SQL slammer attack. The SQL slammer is a worm, which first appeared in 2003. It exploits a buffer overflow vulnerability in the Microsoft SQL Server. When the SQL server receives the request as a single large UDP packet the overrun in the server's buffer leads to the server overwriting its own stack with malicious code. Thereby, the worm code can then be executed. The worm then generates random IP addresses and send itself out to those addresses, allowing to spread rapidly to infect other hosts [7].

2.2 AVOIDIT

Simmons et al. [23] proposed in their paper a cyber-attack taxonomy called AVOIDIT. To classify an attack five classes were used: attack vector, operational impact, defense, informational impact and attack target. In their research they also address the issue of missing consideration of blended attacks in existing taxonomies. A blended attack is an attack which exploits different vulnerabilities at once [22]. So far, only little attention has been given to the possibility of blended attacks. Simmons et al. developed a tree structure for labeling attack vectors in their taxonomy. Their taxonomy is structured in five hierarchical levels. In the following, the classifiers of the first level are introduced. The complete taxonomy can be found in Figure 3 in the appendix of this paper.

Classification by attack vector: An attack vector describes the method or path by which an attacker reaches the target. This classifier defines the vulnerabilities of a system. The attack can use a single attack vector or a combination of several attack vectors. For example, an attacker can use the interaction with users to manipulate them in giving up their confidential information. Thus, social engineering is the way of performing an attack.
Classification by operational impact: This classifier includes the operational effects of an attack. Simmons et al. created a list containing mutual exclusive impacts. When an attacker successfully installs malware e.g. through a script or executable code (see *insufficient input validation* as attack vector), he can gain information about sensitive data.
Classification by informational impact: Besides operational impact the taxonomy also addresses informational impact. Informational impact contains potential ways to effect sensitive information through an attack. Possible impacts are distortion, disruption or disclosure of information.
Classification by attack target: The last classifier defines various attack targets. Possible instances are operating systems, networks, local computers or user information. An attack can also target a combination of instances. **Classification by defense:** The classification by defense contains numerous defense strategies which can be employed before or after an attack occurs. The defence strategies are subdivided in mitigation and remediation. Mitigation covers strategies to diminish damage before or during an attack.

Remediation involves procedures against existing vulnerabilities.

To better understand the process of classifying an attack with AVOIDIT we now demonstrate the usage with the SQL slammer attack. The SQL slammer is a worm launched via installed malware and spreads through the network (**Operational Impact**). It exploits misconfiguration, buffer overflow and denial of service vulnerabilities (**Attack Vector**). Primary targets are networks and applications (**Target**). Several damages can be caused when the worm is successfully installed (**Informational Impact**). It can change access to information (Disrupt), retrieve information (Discover) or modify data (Distort). Preventive and reactive methods are whitelists and patch systems (**Defense**).

2.3 Van Heerden's network attack taxonomy

Van Heerden et al. [27] developed an extensive taxonomy of computer network attacks using 12 classes, each containing multiple sub-classes. Their taxonomy consists of four hierarchical levels. Other than most taxonomies which cover attacks either from an attacker's or defender' point of view, van Heerden et al. included both views in their taxonomy. In the following, a description of the classes of the first level is given. The full taxonomy can be found in Figure 4 in the appendix.

Actor: The actor class describes the different entities which can execute an attack. Subclasses are commercial competitor, hacker, insider or protest groups.
Actor Location: The actor location refers to the country of origin of the attack. Attacks can be launched from local or foreign states. It is also possible that the specific location can not be determined or expand over multiple countries.
Aggressor: The aggressor represents the entity or group launching the attack. Aggressors can be individuals or groups, corporate entities or state aggressors. While the actor class describes the specific type of an attacker, the aggressor is an association with an actor.
Attack Goal: The attack goal specifies the attacker's objective. These can be the breach of security principles such as integrity or availability through changing, destroying or disrupting data. An attack can also work as a springboard for another attack.
Attack Mechanism: The attack mechanism defines the attack methodology. These can be access mechanism like hacking methods, e.g. brute force, phishing and buffer overflow. Data manipulation is another mechanism which uses data as an attack vector. They can be network based, e.g. denial of service or virus-based, e.g. trojans or worms. The collection of information for an attack is classified as information gathering.
Automation Level: This class describes the level of human interaction when launching an attack. A manual attack indicates that an attacker performs the methodology by hand. Automatic attacks only require a minimal amount of input by the attacker. Semi-automatic attacks are launched by tools which require user input.
Effects: Effects describe the severity of consequences caused by an attack. Minor effects are recoverable, whereas major effects are not. Effects are catastrophic when a target can no longer cease as an entity as a result of an attack. However, an attack does not neccessarily have to have an impact on a target. Then it is classified as null.
Motivation: The motivation for an attack differs from aggressor to aggressor. This class specifies incentives for an attack. A common motivation is the financial benefit. Other reasons are criminal or ethical aspects. An Aggressor can also launch attack simply for fun.
Phase: The phase class subdivides an attack into different stages. First, the attacker selects a target. Then the weaknesses of the target are identified. Finally, the attack is executed and post-attack activities are undertaken.
Scope: The scope determines the type of target and its size. Possible types are corporate, governmental or private networks. Corporate and governmetal targets can be subdivided into large or small networks.
Target: This class represents the physical entity targeted by the attack. Targets can be personal computers like laptops and tablets or network infrastructure devices like routers and switches. Servers are other possible targets of an attack.
Vulnerability: The vulnerability class describes the weaknesses exploited by an attacker. These can be deficient configurations regarding access rights or default setups or design issues in protocols or access control. Coding deficiencies are categorizied as implementation vulnerabilities.

Table 1: Classification of the SQL Slammer with van Heerden's [27] taxonomy.

Attack Goal	**Attack Mechanism**	**Automation Level**	**Effect**
Disrupt, Change, Steal	Data-Manipulation: Virus-based: Worm	Automatic	Minor/ Major
Phase	**Scope**	**Target**	**Vulnerability**
Attack	Corporate, Governmental Network	Network, Software	Implementation: Buffer Overflow, Configuration: Default Setup

The resulted classification of the SQL slammer using this taxonomy can be seen in Table 1. The colons represent the hierarchical structure through multiple subclasses. For example the class *Virus-based* is a subclass of the *Data-Manipulation* class and has itself the subclass *Worm*. Actor, Actor Location, Aggressor and Motivation are not listed in the table, since definite values are not available. The effects depend on the target and the severity of the attack. Therefore, effects can be of minor as well as major nature.

3. FROM TAXONOMIES TO ONTOLOGIES

Although taxonomies are useful means for classifications, they lack in several aspects. Taxonomies are often developed for specific domains which makes their extension as well as their consistency problematic. The reuse in other fields is often not possible. While taxonomies have mostly only hierarchical relationships, ontologies can also define custom semantic relationships. The formal and well-structured form of ontologies allow a better communication and reusabil-

doi: 10.2313/NET-2015-03-1_01

ity between organizations [26]. Additional advantages are named in [20]: Ontologies enable the seperation of domain knowledge from operational knowledge. The introduction of relationships provides the possibility to share knowledge with different fields. Ontology languages depict a common information representation and ease the process of information reuse.

According to Noy and McGuiness [20] an ontology consists of **concepts**, **attributes of classes** and **restrictions of slots**. The concept of a domain is described by classes. A class can have multiple subclasses which describe more specific concepts. Each class or subclass has instances. For example is *food* a superclass, *vegetable* and *fruit* subclasses and *apple* and *broccoli* are instances. The arrangement of the classes in a hierarchy builds the underlying taxonomy of the ontology. First-level classes are also referred as concepts. Attributes of classes are called slots. They describe behavioral and semantic properties of classes. Slots can therefore be described as relationships between classes. The class *human* for example can have the subclasses *woman* and *man*. Between those two subclasses a relationship can be defined, e.g. a man *is a husband* to a woman. Finally, ontologies need to define restrictions of slots, also called facets. Facets describe allowed values or types a slot can take. In the example above possible restrictions would be that a woman can have 0 or 1 husband, but not more.
The need of an ontology has been identified and there have been various attempts to create security ontologies [26, 14, 9]. In the following, two security ontologies are introduced in more detail.

3.1 Van Heerden's Ontology

In the previous section the taxonomy of van Heerden et al. was presented. This taxonomy is now used to create an ontology. The definition and arrangement of the classes are realized in their taxonomy. Furthermore, for their ontology they added an "Attack Scenario" class. This class is used to classify computer attacks and connects the other classes. It is subdivided in the classes denial of service, industrial espionage, web deface, spear phishing, password harvesting, snooping for secrets, financial theft, amassing computer resources, industrial sabotage and cyber warfare. Every attack scenario has a scope and goal. It consists of different attack phases and is assigned to an actor and an aggressor.
The next step in the ontology development process is to define the slots. Every class and subclass has a *is-a*-relationship. Classes can also have inter-relationships. Every *Actor* has at least one *Actor Location*. An *Aggressor* has always a motivation. An *Attack Mechanism* has exactly one *Target* and one *Automation Level*. A *Phase* has one *Effect* and a *Attack Mechanism*. A *Target* has a *Vulnerability* and a *Attack Scenario* has a *Attack Goal*, a *Phase* and at least one *Actor* and *Aggressor*. All relationships can be seen in Figure 1. A rectangle represents a class, the arrow the *has*-relationship between classes.

For the subclasses of the Attack Scenario class van Heerden et al. additionally defined attribute restrictions for their slots. With those restrictions attacks can be clearly seperated from each other. However, as they state themselves in their publication, their list of attack scenarios does not cover the full scope of possible attacks.

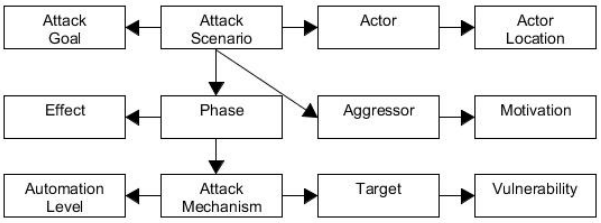


Figure 1: Van Heerden et al.'s [27] ontology

If the ontology is applied to the example of the SQL Slammer, following descriptions are defined: The Attack Scenario is SQL Slammer. It has the Phase attack. It has the Attack Mechanism worm. The SQL Slammer targets networks and software utilizing the vulnerabilities buffer overflow and default setup. Goal is to disrupt, change or steal data.

3.2 Ontology-based attack model

Gao et al. [10] developed an ontology-based attack model to assess the security of an information system from the angle of an attacker. Goal of the assessment process is the evaluation of attack effects. Thereby, the difference of system performance before and after an attack is calculated. The process consists of four phases. First, vulnerabilities of the system are identified using automated vulnerability tools. Such tools assess computer system, applications or network regarding their vulnerabilities and generate sets of scan results. In the second phase, the developed ontology is used to determine which attacks might occur due to the identified vulnerabilities. By quering the ontology, the possible effects are optained. This is the third phase. Finally, in the last phase the attack effect is calculated. In this paper a short overview of the classes is provided. For more detailed insight the reader is referred to their publication.

The ontology of [10] holds five classes: attack impact, attack vector, attack target, vulnerability and defense. **Attack Impact** consists of the security principles confidentiality, integrity, availability, authentication, authorization and auditing. All these principles are security properties of the target threatened by an attack. The **Attack Vector** describes here also the path by which am attack is launched. The **Target** class contains the possible targets hardware, software and humans. The **Vulnerability** addresses weaknesses and defects of the system. These can be for example design or implementation flaws. Finally, the **Defense** class describes countermeasures against attacks. The classes of their ontology show similarities to those used in the AVOIDIT taxonomy. Both adopted concepts from [13], [15] and [14].

Gao et al. [10] used relationships defined by Herzog [14] and extended his definitions with additional relationships. An attack *has* one or more attack vectors. It is *enabled by* a vulnerability. An attack *threatens* security properties defined in the attack impact. An attack vector *threatens* a target which *has* vulnerabilities. A target can also *reside* in another target. Defense strategies *protect* the target and the security properties. Finally, relationships between attack vectors are realized with the *ifSuccessfulLeadsToThreat*-relation. The ontology with all relations is shown in Figure 2.

doi: 10.2313/NET-2015-03-1_01

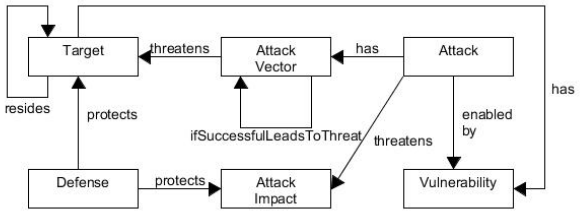


Figure 2: Gao et al.'s ontology [10]

Now the SQL Slammer is applied to the ontology. The SQL Slammer is a computer worm and *has* the attack vectors buffer overflow and denial of service. The attack *is enabled* by the vulnerabilities due to implementation flaws. Threatened targets are networks. If a Slammer attack succeeds he can cause further DoS attacks.

4. ANALYSIS AND RESULTS

After examining several classifications, a comparative analysis is conducted in this section. Thereby, the taxonomies and ontologies are compared with each other. Later on, the taxonomies are compared to the ontologies to determine the advantages of tansitioning from taxonomies to ontologies.

4.1 Taxonomies

The comparison of taxonomies is not a straightforward task. No general methods to compare attack taxonomies have been proposed so far. For the comparison of the presented taxonomies the criteria for developing successful taxonomies presented in Section 2.1 are applied.
First, a general comparison between the defined classes in the taxonomies of van Heerden et al.'s work and the AVOIDIT taxonomy is made.

Both taxonomies define a target class for possible targets. While van Heerden et al. provides a deeper hierarchical order of the target class using three subclasses, AVOIDIT uses a wider partion of targets with six subclasses. Van Heerden et al. provide moreover the scope class, which can be seen as an addition to the target class, giving more detail about the size and type of the target.
The classes attack goal in van Heerden et al.'s taxonomy and informational impact represent both the purpose of an attack. They share the same subclasses change, destroy and disrupt data. AVOIDIT provides beyond that two more subclasses disclosure and discover for acquiring information. Van Heerden et al. limits this to the single subclass steal data. The classes vulnerability and attack vector cover both security flaws and weaknesses that build the path to a successful attack. Van Heerden et al. provides excessive information about the attacker with additional classes, while AVOIDIT does not cover this aspect. This is due to fact, that in contrary to AVOIDIT, van Heerden et al.'s taxonomy not only addresses the defenders's point of view, but also the attackers's. Therefore, additional information about the attacker including location and motivation is neccessary. Furthermore, they provide additional information about the attack describing the different phases as well as the automation level of an attack. Other than van Heerden et al., AVOIDIT moreover provides defense techniques against attacks.

Table 2: Comparison of complied taxonomy requirements

Requirements	van Heerden et al.	Simmons et al.
Accepted	y	y
Comprehensible	y	y
Conforming	y	y
Determined	y	y
Exhaustive	n	n
Mutual Exclusive	y	y
Repeatable	y	y
Well Defined	n	y
Unambiguous	y	y
Useful	y	y

Now both taxonomies are evaluated against the criteria for a sufficient taxonomy. An overview of the comparison can be seen in Table 2. Van Heerden et al. state in their paper, that their taxonomy does not fulfill all criteria. Completeness could not be achieved due to the wide scope of existing attacks. Because their ontology uses a rather wide definition of network attacks instead of a detailed definition, also the requirement of well-defined terms was not achieved. According to the authors their developed taxonomy complies in the remaining requirements.

The AVOIDIT taxonomies meets all criteria for sufficient taxonomies according to their authors. However, since constantly new attacks and vulerabilities approach, their taxonomy is not considered exhaustive. Furthermore, the criteria for determinism is not mention in their publication. Since a detailed description about the development of their classes is provided, their taxonomy is considered determined.

Both authors name limitations of their developed taxonomy. Van Heerden et al.'s taxonomy does not cover all possible attack scenarios. AVOIDIT on the other hand, lacks in the amount of defense strategies. Both taxonomies do not discuss physical attacks.

4.2 Ontologies

In this section, the presented ontologies are analyzed. Thereby, first a general comparison is conducted analyzing differences and similarities between concepts, classes, slots and facets of each ontology. Based on the work of [5], we construct a table containting comparative metrics such as number of classes, average number of slots and average number of subclasses. Finally, we conclude by analyzing limitations and neccessary future work.

Ontologies consist of classes which are hierarchically ordered in a taxonomy. Van Heerden et al's taxonomy consists of overall 12 classes. In addition, for their ontology they added another class, the Attack Scenario, which makes their ontology consist of overall 13 classes. The ontology of Gao et al. contains six different classes.
Every taxonomy realizes the *is-a*-relationship between classes and subclasses. In an ontology, further relationships or slots can be defined. For their ontology, van Heerden et al. de-

doi: 10.2313/NET-2015-03-1_01

Table 3: Comparison of general metrics between [27] and [10]

Metric	van Heerden et al.	Gao et al.
Number of concepts	13	6
Avg. number of subclasses/concept	3.6	8.2
Avg. depth of inheritance	2.8	2.2
Number of slots	10	9
Avg. number of slots/concept	1.8	2.6

fined the *has*-relationship to represent inter-relationships between classes. Gao et al.' taxonomy consists of a broader range of relationships.
To make the ontology complete van Heerden et al. define several restrictions of slots. These restrictions help to clearly distinct between attacks. Therefore, they define ten different attack scenarios with unique constraints. Gao et al. depict constraints for the three attacks SQL Slammer, Rootkit and the Mitnick attack.

Now the ontologies are compared using the metrics stated in [5]. The results are displayed in Table 3. The findings show, while van Heerden et al. use more concepts, Gao et al. have more subclasses per concept. The calculation of the average number of subclasses only includes subclasses until the second level. The average depth of inheritance describes the number of hierarchical levels for every concepts. Van Heerden et al. define their concepts in greater depth than Gao et al.. Regarding the relations between concepts, both define almost equal number of slots. However, Gao et al. have defined more slots per concept than van Heerden et al. in their ontology.

4.3 Taxonomies vs. Ontologies

So far, taxonomies have not directly been compared to ontologies. To determine the differences and similarities, the presented taxonomies and ontologies are compared with each other. Thereby, we compare the following aspects: purpose, usage, relationships and representation. The results will give further insights into the categorization process of an attack with the different concepts. Furthermore, we will conclude with advantages and disadvantages depending on the results of the comparison.

Purpose: Both taxonomies and ontologies follow the purpose to index attacks by classifying them by their characteristics. The AVOIDIT taxonomy is used to provide information regarding attack vectors, possible effects and defense strategies about an attack. Besides indexing attacks the ontologies describe a domain of knowledge. Van Heerden et al.'s taxonomy and ontology is supposed to clearly classify an attack from the view of the attacker and the target. As a future task they mention the refinement of their ontology to apply it for attack prediction. The main purpose of Gao et al.'s ontology is the evaluation of an attack effect.

Usage: The AVOIDIT taxonomy is applied to an issue resolution system (IRS). The IRS is a system which contains and manages a list of issues and countermeasures for those issues. It teaches the defender about potential risks of cyber attacks. The list is organized according to the taxonomy. Their taxonomy does not provide any information if an attack was successful, but classifies the attack vectors to foresee possible effects and identify appropriate defense strategies. Van Heerden et al. do not state any specific information on where their taxonomy and ontology is applied. In their future work they mention the usage of their ontology in intrusion detection systems. However, the determination of concepts such as motivation or attack goal by a computer system seems to be problematic. Gao et al. use their ontology in an ontology-based framework. The framework calculates the attack effect by comparing the system performance before and after the attack.

Relationships: Due to the hierarchical structure a taxonomy can only provide a parent-child-relationship. Ontologies, however, can not only describe a domain in a hierarchy but also define additional relations between classes and different concepts. These relationships are of a semantical or behavioral character. The AVOIDIT taxonomy for example allows the relations *is-a* between the target and its subclasses. The ontology by Gao et al. additionally adds the relation *resides* between different targets. This provides relationships between different concepts. Therefore, a taxonomy can be seen as a tree, whereas an ontology functions more like a web. This concludes that taxonomies are often restricted to the usage in a specific domain. Ontologies on the other hand allow the communication to other concepts and systems. This also points to another restriction of taxonomies, namely that knowledge is in most cases not hierarchical.

Representation: Taxonomies are mostly represented graphically in a tree-like structure. Ontologies can be represented either in a formal text format or graphically. Through machine interpretable definitions of the concepts computer applications are capable of interpreting the ontology. Gao et al.'s ontology is build using the language OWL. OWL is based on XML and is endorsed by the World Web Consortium (W3C). Van Heerden et al. make no further statements regarding the language they used for their ontology. Both presented taxonomies use a tree structure for their realization. The advantage of ontologies over taxonomies in this aspect is that the use of machine interpretable definitions makes reusability and knowledge sharing between different software systems easier.

The purpose of a taxonomy is to provide useful means to classify characteristics of attacks and thereby provide a better description of attacks. This classification helps to identify vulnerabilities, predict potential attacks and possible effects. Taxonomies are mostly used for risk management with identification, assessment and prioritization of risks as well as evaluation of systems. Taxonomies do not determine if an attack was successful. The AVOIDIT taxonomy is used in an issue resolution system. It classifies the attack vector information and foresees possible effects on the system. In summary, taxonomies are primarily used to represent security knowledge and determine defense mechanisms prior an attack.

Ontologies, unlike taxonomies, use semantic relations between attacks. Machine interpretable syntax allows comprehensive use in software systems such as Intrusion Detection Systems. Monitoring components collect data such as traffic, requests or packets and an alerting system provides response on the attempted attack and countermeasures. Gao et al. use their ontology for security assessment. Thereby, first vulnerabilities of the system are detected. Then possible attacks are queried. Based on the resulting attacks risks and neccessary defense methods are determined.

5. RELATED WORK

The use of taxonomies has become a key technique for the categorization and formal description of attacks. They reach from general attack taxonomies to specific field related taxonomies. Until today numerous surveys were conducted analyzing existing taxonomies to use them in defense methods against network attacks. Igure and Williams [16] conducted an extensive survey on cyber adversaries and attacks, discussing taxonomies from the early 1970s to 2006. By analyzing the efficiency of these taxonomies regarding the use in security assessment, Igure and Williams define requirements for taxonomies used in a security assessment process. Another extensive survey was presented by Meyers et al. [21]. In their paper, publications from 1985 to 2006 are covered. Further surveys were carried out in [29], [25] and [13].

Although many surveys were conducted on existing taxonomies, only few research was done regarding attack ontologies. Blanco et al. [3] carried out a systematic survey on existing security ontologies, evaluating and comparing concepts, relations and attributes using a framework. Souag et al. [24] conducted a general survey on existing security ontologies. Furthermore, the examined ontologies were analyzed regarding security aspects such as vulnerabilities, threats and countermeasures and evaluated for the use in security requirements engineering. Evesti et al. [8] examine a number of security ontologies, comparing their applicability for run-time security monitoring.

6. CONCLUSION

In this paper, a survey on existing attack taxonomies and ontologies was conducted. Furthermore, an analyis comparing differences between those concepts was carried out.
While many taxonomies for specific fields exist, there has been an increased attempt to develop a common, standardized attack taxonomy for the scientific community. However, depending on their goal and purpose the taxonomies still differ in their realization. Furthermore, the development of most examined work still resides in the early stages since they do not completely cover all attack possibilities. Therefore, the neccessity to combine existing taxonomies was identified.
Like research before has already shown [26] the limitations of attack taxonomies make the advancement to ontologies a neccessary task. The development of ontologies has been identified as an important branch of research. As a result of this work, it is concluded that the existing taxonomies and ontologies are not far enough developed for general usage and extension. Existing concepts need to be combined to create a flexible ontology that easily enables reuse and knowledge sharing between different systems.

7. REFERENCES

[1] E. G. Amoroso: *Fundamentals of Computer Security Technology* Prentice-Hall PTR, 1994

[2] M. Bishop, D. Bailey: *A critical analysis of vulnerability taxonomies*, California University Davis, Department of Computer Science, 1996

[3] C. Blanco, J. Lasheras, R. Valencia-Garcia, E. FernÃąndez-Medina, A. Toval, M. Piattini: *A systematic review and comparison of security ontologies*, In Availability, Reliability and Security, pages 813-820, ARES 08. Third International Conference on, 2008

[4] E. Blomqvist, A. Ohgren, K. Sandkuhl: *The analytic hierarchy process and multicriterion decision making* In: Enterprise Information Systems, pages 221-240, Springer Berlin Heidelberg, 2008

[5] E.Blomqvist, A. Ohgren, K. Sandkuhl: *Ontology Construction in anEnterprise Context: Comparing and Evaluating Two Approaches* In: Proceedings of the Eighth International Conference on Enterprise Information Systems: Databases and Information Systems Integration,Paphos, Cyprus, 2006

[6] K. F. P. Chan, M. Olivier, R.P. van Heerden: *A Taxonomy of Web Service Attacks*, In: Proceedings of the 8th International Conference on Information Warfare and Security: ICIW 2013, page 34, Academic Conferences Limited, 2013

[7] T. M. Chen, J. M. Robert(2004): *Worm epidemics in high-speed networks.* Computer, 37(6), pages 48-53, 2004

[8] A. Evesti, E. Ovaska, R. Savola: *From security modelling to run-time security monitoring*, Security in Model-Driven Architecture, page 33, 2009

[9] S. Fenz, A. Ekelhart: *Formalizing information security knowledge* In: Proceedings of the 4th international Symposium on information, Computer, and Communications Security, ACM, pages 183-194, 2009

[10] J. B. Gao, B. W. Zhang, X. H. Chen, Z. Luo: *Ontology-based model of network and computer attacks for security assessment* Journal of Shanghai Jiaotong University (Science), 18. Jg., pages 554-562, 2013

[11] T. R. Gruber: *A translation approach to portable ontology specifications*,Knowledge acquisition, 5. Jg., Nr. 2, pages 199-220, 1993

[12] N. Gruschka, M. Jensen: *Attack surfaces: A taxonomy for attacks on cloud services*, In: Cloud Computing (CLOUD), 2010 IEEE 3rd International Conference on. IEEE, pages 276-279, 2010

[13] Hansman, Simon, and R. Hunt: *A taxonomy of network and computer attack methodologies*, Department of Computer Science and Software Engineering, University of Canterbury, 7, New Zealand, 2003

[14] A. Herzog, N. Shahmehri, C. Duma: *An ontology of information security* International Journal of Information Security and Privacy (IJISP), 1. Jg., Nr. 4, Spages. 1-23, 2007

[15] J. D. Howard: *An Analysis Of Security Incidents On The Internet 1989-1995* PhD thesis, Carnegie Mellon University, 1997.

[16] V. Igure, R. Williams: *Taxonomies of attacks and*

vulnerabilities in computer systems, Communications Surveys & Tutorials, IEEE, 10, 1, pages 6-19, 2008

[17] I. V. Krsul: *Software Vulnerability Analysis* PhD thesis, Purdue University, 1998

[18] C. E. Landwehr, A. R. Bull, , J. P. McDermott, W. S. Choi: *A taxonomy of computer program security flaws, with examples*, Naval Research Lab Washington DC, 1993

[19] U. Lindqvist, E. Jonsson: *How to Systematically Classify Computer Security Intrusions* IEEE Security and Privacy, pages 154âĂŞ163, 1997

[20] N. F. Noy, D. L. McGuinness: *Ontology development 101: A guide to creating your first ontology* Stanford University, Stanford, CA, 2001

[21] C. Meyers, S. Powers, D. Faissol: *Taxonomies of cyber adversaries and attacks: a survey of incidents and approaches*, Lawrence Livermore National Laboratory (April 2009), 7, 2009

[22] D. Moore, V. Paxson, S. Savage, C. Shannon, S. Staniford, N. Weaver: *Inside the slammer worm*In IEEE Security and Privacy, volume 1, 2003.

[23] C. Simmons, C. Ellis, S. Shiva, D. Dasgupta, Q. Wu: *AVOIDIT: A cyber attack taxonomy*, Annual Symposium on Information Assurance, 2014

[24] A. Souag, C. Salinesi, I. Comyn-Wattiau: *Ontologies for security requirements: A literature survey and classification*, In: Advanced Information Systems Engineering Workshops. Springer Berlin Heidelberg, pages 61-69, 2012

[25] M. Uma, G. Padmavathi: *A Survey on Various Cyber Attacks and their Classification*, IJ Network Security, 15(5), pages 390-396, 2013

[26] J. Undercoffer, A. Joshi, J. Pinkston: *Modeling computer attacks: An ontology for intrusion detection*, In: Recent Advances in Intrusion Detection. Springer Berlin Heidelberg, pages 113-135, 2003

[27] R. P. van Heerden, B. Irwin, I. D. Burke: *Classifying network attack scenarios using an Ontology*, In: Proceedings of the 7th International Conference on Information Warfare and Security. Academic Conferences Limited, pages 331-324, 2012

[28] L. G. Vargas, J.J. Doughe: *The analytic hierarchy process and multicriterion decision making* American Journal of Mathematical and Management Sciences, , 19(1), pages 59-92, 1982

[29] J. Wei: *Survey of network and computer attack taxonomy*, Proceedings of the 2012 IEEE Symposium on Robotics and Applications (ISRA), IEEE, USA, 2012

APPENDIX

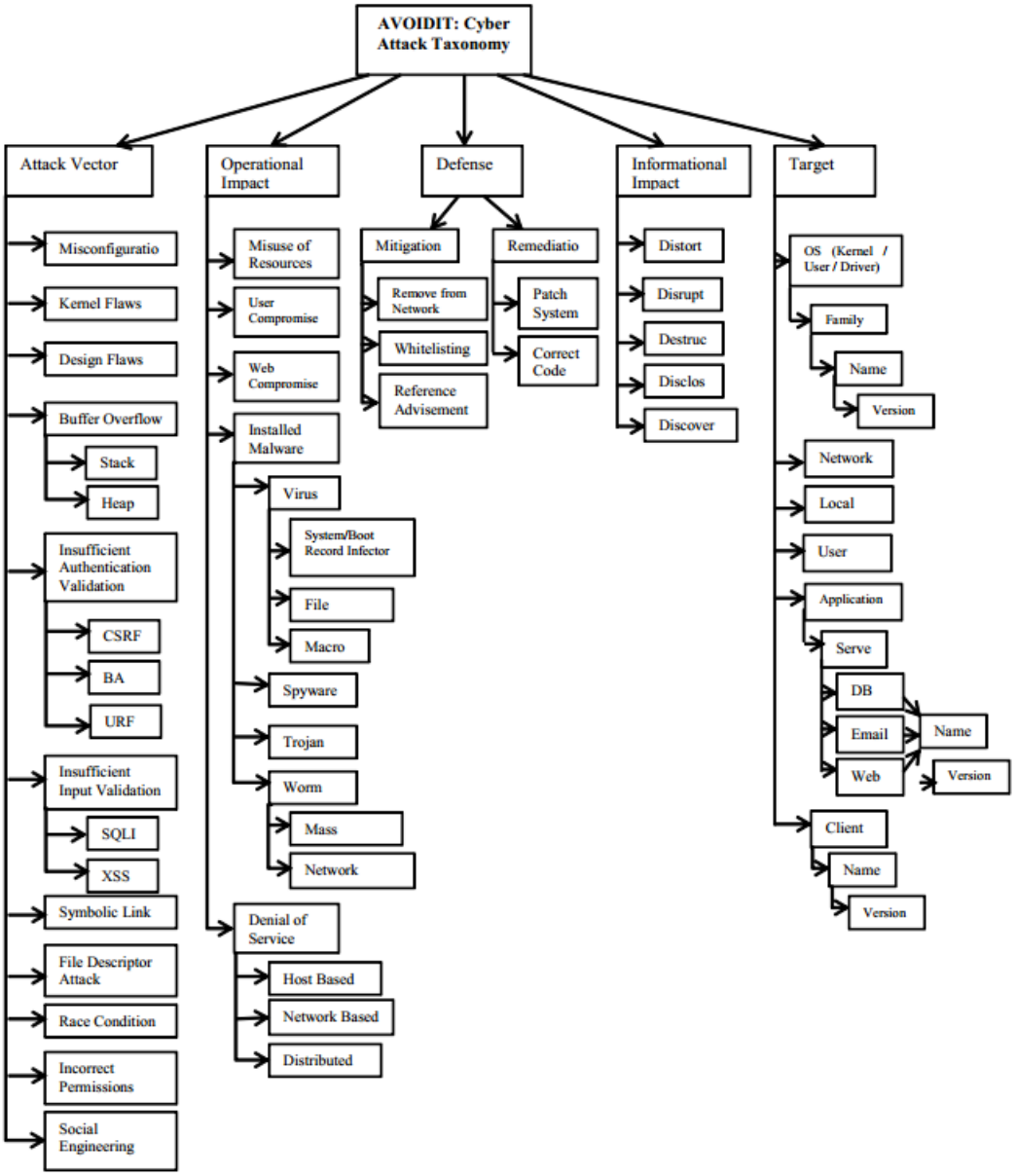


Figure 3: AVOIDIT taxonomy [23]

doi: 10.2313/NET-2015-03-1_01

1. Actor 1.1 Commercial Competitor 1.2 Hacker 1.2.1 Script Kiddie Hacker 1.2.2 Skilled Hacker 1.3 Insider 1.3.1 Admin Insider 1.3.2 Normal Insider 1.4 Organised Criminal Group 1.5 Protest Group	2. Actor Location 2.1 Foreign Actor Location 2.2 Local Actor Location Indeterminate Actor Location
3. Aggressor 3.1 Individual Aggressor 3.2 Commercial Aggressor 3.3 State Aggressor 3.4 Group Aggressor 3.4.1 Ad-hoc Group Aggressor 3.4.2 Organized Group Aggressor	4. Attack Goal 4.1 Change Data 4.2 Destroy Data 4.3 Disrupt Data 4.4 Steal Data Springboard for other attack goal
5. Attack Mechanism 5.1 Access 5.1.1 Brute Force 5.1.2 Buffer Overflow 5.1.3 Spear Phishing 5.1.4 Physical 5.2 Data Manipulate 5.2.1 Network-based 5.2.1.1 Denial of Service 5.2.2 Virus-based 5.2.2.1 Trojan 5.2.2.2 Virus 5.2.2.3 Worm 5.2.3 Web-Application-based 5.2.3.1 SQL Injection 5.2.3.2 Cross-site scripting 5.3 Information Gathering 5.3.1 Scanning 5.3.2 Physical	6. Vulnerability 6.1 Configuration 6.1.1 Access Rights 6.1.2 Default Setup 6.2 Design 6.2.1 Open Access 6.2.2 Protocol Error 6.3 Implementation 6.3.1 Buffer Overflow 6.3.2 Race Condition 6.3.3 SQL Injection 6.3.4 Variable Type Checking
7. Effects 7.1 Null 7.2 Minor Damage 7.3 Major Damage 7.4 Catastrophic	8. Motivation 8.1 Financial 8.2 Fun 8.3 Ethical 8.4 Criminal
9. Phase 9.1 Target Identification 9.2 Reconnaissance 9.3 Attack Phase 9.3.1 Ramp-up 9.3.2 Damage 9.3.3 Residue 9.4 Post- Attack Reconnaissance	10. Scope 10.1Corporate Network 10.1.1 Large Corporate Network 10.1.2 Small Corporate Network 10.2Government Network 10.2.1 Large Government Network 10.2.2 Small Government Network 10.3Private Network
11. Target 11.1Personal Computer 11.2Network Infrastructure Device 11.3Server	12. Automation Level 12.1Manual 12.2Automatic Semi-Automatic

Figure 4: Van Heerden et al.'s taxonomy [27]

Freenet

Florian Baumann
Betreuer: Dr. Holger Kinkelin, Marcel von Maltitz
Seminar Future Internet WS2014
Lehrstuhl Netzarchitekturen und Netzdienste
Fakultät für Informatik, Technische Universität München
Email: derflob@mytum.de

KURZFASSUNG

Das Internet ist aufgrund seiner Architektur und heutigen Implementierung anfällig gegenüber Zensur durch mächtige Parteien wie beispielsweise Regierungen. Zudem bietet es kaum Möglichkeiten für das anonyme Speichern und Abrufen von Daten. *Freenet* hingegen bietet eine anonyme, zensurresistente Peer-to-Peer-Infrastruktur zum Speichern und Aufrufen von Daten. Bei der Konzeption der von *Freenet* verwendeten Algorithmen wurde Wert auf eine effiziente Funktionsweise gelegt. Dazu wurde sich an sozialen Strukturen unserer Gesellschaft orientiert, die sich auch im *Freenet* widerspiegeln.

Schlüsselworte

Anonymität, Peer-to-Peer-Netzwerk, Verteilte Datenspeicherung, Zensurresistenz

1. EINLEITUNG

Im Internet gespeicherte Daten können zensiert werden. Dies ist darin begründet, dass die Herkunft von Daten immer einem Server zuordenbar ist. So können beispielsweise Regierungen durch Beschlagnahmung entsprechender Server die darauf enthaltenen Daten unerreichbar machen und damit zensieren. Es sind außerdem noch andere Arten von Attacken, beispielsweise auf das Domain Name System (DNS) denkbar, bei denen ein Angreifer, der globalen Zugriff und Manipulationsmöglichkeiten im Netzwerk hat, DNS-Anfragen umleitet. Da Kommunikation auch heute noch zu einem Großteil unverschlüsselt stattfindet, ist es solchen Angreifern auch möglich die Konsumenten und Produzenten von Information zu überwachen und deanonymisieren. Um dennoch einen anonymen und zensurfreien, beziehungsweise -resistenten Informationsaustausch zu gewährleisten, sind zusätzliche Maßnahmen vonnöten, die auf den bestehenden Netzwerken aufbauen.

So existiert eine Vielzahl von Anonymisierungsanwendungen, wie etwa *TOR* (The Onion Router)[1], diverse VPN-Dienste, *I2P*[2], *GNUnet*[3] oder *Freenet*[4]. In dieser Arbeit wird *Freenet* betrachtet, welches ein System zum anonymen, verteilten Speichern und Aufrufen von Daten ist. Es wurde 1999 von *Ian Clarke* vorgestellt[1] und wird seitdem ständig erweitert und verbessert. Zur Wahrung der Anonymität der Benutzer im *Freenet* findet einerseits eine Transportverschlüsselung zwischen den einzelnen Peers statt, so dass es einem Eavesdropper nicht möglich ist die Kommunikation passiv abzuhören. Andererseits wird die Herkunft von Anfragen und Daten als Teil des Routing-Protokolls verschleiert, sodass die Quelle nicht eindeutig identifiziert werden kann. Mit der Version 0.7 aus dem Jahr 2005 wurde das verwendete Routing-Protokoll grundlegend verändert, um das Auffinden von Daten effizienter zu gestalten. Gleichzeitig wurde es möglich, *Freenet* in einem *Darknet*-Modus zu verwenden, um die Sicherheit und Anonymität weiter zu erhöhen.

Im Kapitel 2 wird darauf eingegangen, welche Funktionen und Sicherheitsversprechen *Freenet* bietet, sowie worum es sich bei einem *Darknet* handelt und wie es in *Freenet* integriert ist. Kapitel 3 erläutert die verschiedenen Schlüsseltypen, die beim Routing zum Einsatz kommen und gibt einen Überblick, wie Verschlüsselung eingesetzt wird um Benutzer zu schützen und die Sicherheitsversprechen aus Kapitel 2 umzusetzen. Das Kapitel 4 beschreibt die Algorithmen, die für das Routing innerhalb des *Freenet* benutzt werden. In Kapitel 5 wird ein kompakter Überblick gegeben, welche zusätzlichen Diensttypen durch Plugins oder Third-Party Programme implementiert wurden und auf den Konzepten von *Freenet* aufbauen. Das Kapitel 6 zeigt einige Probleme auf, unter denen *Freenet* noch leidet und präsentiert etwaige Lösungsüberlegungen. Abschließend wird in Kapitel 7 ein Überblick über verwandte Arbeiten gegeben, bevor in Kapitel 8 eine Zusammenfassung folgt.

2. ZIELE UND EIGENSCHAFTEN

Freenet implementiert eine Art verteiltes Peer-to-Peer-Dateisystem als Overlay-Netzwerk auf dem heutigen Internet. In der ersten Veröffentlichung zum *Freenet* wurden die folgenden Eigenschaften definiert, die dieses Overlay-Netzwerk haben soll[1]:

Dezentralisierung Um keinen Angriffspunkt für (Distributed) Denial of Service Attacken (DDoS) zu bieten, die zum Ziel haben können Information zu unterdrücken, soll das System komplett ohne zentrale Verwaltungs- oder Kontrollmechanismen auskommen. Dies steht zum Beispiel im Gegensatz zu *Napster*, welches zwar eine Peer-to-Peer-Anwendung war, jedoch einen zentralen Server besaß, welcher für die Koordinierung und Vermittlung der Downloads zuständig war.

[1] `https://www.torproject.org/`
[2] `https://geti2p.net/`
[3] `https://gnunet.org/`
[4] `https://www.freenetproject.org/`

doi: 10.2313/NET-2015-03-1_02

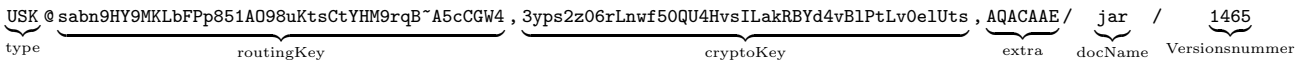


Abbildung 1: Beispiel für einen Updateable Subspace Key, URI für *Freenet* -Updater

Anonymität Sowohl Entitäten, die Information anbieten, als auch jene, die diese aufrufen sollen anonym bleiben.

Robustheit Software- und Hardwarefehler und der damit verbundene Ausfall von Peers sollen das System nicht stark beeinflussen können.

Adaptivität Aufgrund der Natur von P2P-Netzwerken, bei denen Peers ständig verschwinden und Neue beitreten, soll *Freenet* in der Lage sein, sich schnell und effektiv an die neuen Gegebenheiten anpassen zu können.

Performance Die Geschwindigkeit soll bestehenden Informationssystemen gleichen.

Diese Eigenschaften ergeben sich unter anderem aus dem Fokus auf eine Widerstandsfähigkeit gegen Zensur durch Regierungen, die direkten Zugriff und Manipulationsmöglichkeiten auf die bestehende Infrastruktur haben können.

Ein weiteres Ziel ist eine **Plausible Deniability** darüber, was auf dem eigenen Rechner gespeichert ist, um eine Verfolgung von Teilnehmern zu erschweren oder zu verhinden.

Eine wichtige Eigenschaft im Zusammenhang mit Datenspeichern, eine **Lebenszeitgarantie** von Dokumenten, verspricht *Freenet* explizit nicht. Stattdessen wird argumentiert, dass Informationen, die von Interesse sind, deshalb „überleben“, weil sie regelmäßig aufgerufen werden.

Daten werden zum Routing und Auffinden mit einem global eindeutigen Schlüssel versehen. Es sind mehrere Schlüsseltypen in Verwendung, die in Kapitel 3 genauer betrachtet werden. Jeder Peer besitzt zudem eine *Location*, die durch eine Fließkommazahl im Bereich von 0 bis 1 ausgedrückt wird. Sie hat keinen Zusammenhang mit der geographischen Position des Peers, sondern wird beim Beitreten in das Netzwerk zufällig erstellt und dient nur dem Routing. Die Locations der Nachbarn, mit denen ein Node verbunden ist, werden lokal in einer Routing-Tabelle gespeichert.

Freenet ist zudem mit Version 0.7 von einem TCP- auf einen reinen UDP-Transport umgestiegen. Dabei wurde jedoch nachträglich, aufbauend auf UDP, eine Congestion Control und eine zuverlässige Packetvermittlung implementiert.

2.1 Darknet & Opennet

Mit der Veröffentlichung der Version 0.7 von *Freenet* hat das Konzept des *Darknet* Einzug gehalten. Wird *Freenet* im Darknet-Modus betrieben, werden nur Verbindungen zu solchen Peers hergestellt, die man manuell als „Freunde“ eingetragen hat. Dazu müssen beide Parteien ihre *Node References* austauschen, bei welchen es sich um diverse Parameter – wie etwa IP-Adresse oder kryptographische Parameter – handelt, die notwendig sind, eine sichere, verschlüsselte Verbindung aufzubauen. So entsteht nach und nach ein vertrauenswürdiges Netzwerk, das reale, soziale Netzwerke widerspiegelt. Da beim Routing nur Anfragen an direkt verbundene Nachbarn geschickt werden, denen man vertraut die eigene Anonymität nicht zu kompromittieren, erhöht sich die Anonymität und Sicherheit gegenüber böswilligen Angreifern.

Da ein Darknet jedoch eine erhebliche Einstiegshürde darstellt, bietet *Freenet* weiterhin einen Opennet-Modus an. Dabei werden Verbindungen zu Dritten, die nicht explizit vom Benutzer hinzugefügt wurden, sondern im Laufe der Zeit entdeckt wurden, erlaubt und hergestellt.

Beide Modi lassen sich in Form eines hybriden Modus kombinieren, bei dem Peers aus dem Darknet bevorzugt behandelt werden und nur dann Verbindungen in das Opennet hergestellt werden, falls das Limit für die maximale Anzahl an Verbindungen noch nicht erreicht wurde.

3. PROTOKOLLGRUNDLAGEN

3.1 Schlüsseltypen: CHK, SSK, USK

Es werden hauptsächlich drei verschiedene Typen von Schlüsseln verwendet um Daten aufzufinden. *Content Hash Keys* (CHK), *Signed Subspace Keys* (SSK) und *Updateable Subspace Keys* (USK). Abbildung 1 stellt beispielhaft einen USK mit seinen einzelnen Bestandteilen dar. Bis auf das Versionsnummernfeld, das für USKs einzigartig ist, bestehen alle Schlüsseltypen aus den gleichen Feldern, die in Abhängigkeit des Typs leicht unterschiedliche Bedeutungen haben. Dabei sind die Felder *routingKey*, *cryptoKey* und *extra* Base64 kodiert, mit „-“ und „~“ anstelle der sonst üblichen Zeichen „/“ und „+“ um die URIs als URLs in einem Browser verwenden zu können. Das extra-Feld gibt zusätzliche Parameter, wie etwa den Verschlüsselungsalgorithmus an. *docName* ist der menschenlesbare Dateiname der hochgeladenen Datei.

Der grundlegendste Schlüsseltyp ist der CHK. Er ist vor allem dafür geeignet, große, statische Daten in das *Freenet* einzufügen. Der routingKey ist dabei der SHA-256-Hash über die eingebrachten Daten. Dadurch entsteht für jedes Datum ein global eindeutiger Identifier, da Hashkollisionen, gezielt oder zufällig, als überaus unwahrscheinlich gelten.

SSKs bieten mit Hilfe asymmetrischer Kryptographie eine Möglichkeit private Namespaces im *Freenet* zu schaffen. Nur der Besitzer des privaten Schlüsselteils eines Subspace ist in der Lage Daten hinzuzufügen. Dazu wird zum Erstellen eines neuen Subspaces ein zufälliges Schlüsselpaar generiert. Als routingKey wird der Hash des öffentlichen Schlüsselteils verwendet, der damit gleichzeitig den Subspace identifiziert. Soll nun eine Datei zu einem Subspace hinzugefügt werden, muss der Identifier dieser Daten berechnet werden. Dazu wird der docName gehasht, mit dem Hash des öffentlichen Schlüssels XOR-verknüpft und erneut gehasht. Bevor ein Subspace-Verwalter nun Daten in das *Freenet* hochlädt, fügt er den Daten noch eine Signatur, die nur er mit dem

doi: 10.2313/NET-2015-03-1_02

privaten Schlüssel erstellen kann sowie zusätzlich auch den öffentlichen Schlüssel an. Da mit der Signatur die Authentizität der bereits verschlüsselten Daten gewährleistet wird, kann jeder Node der ein Datum sieht die Signatur auf ihre Gültigkeit überprüfen, ohne den echten Dateninhalt kennen zu müssen.

Versucht nun ein Angreifer Daten in ein Subspace einzufügen, für das er den privaten Schlüssel nicht kennt, ist er nicht in der Lage eine gültige Signatur zu erstellen und die Daten werden von den anderen Nodes abgewiesen.

USKs sind im Grunde SSKs, nur bieten sie eine automatische Möglichkeit der Versionierung von Daten unter dem gleichen Namen. Da bei SSKs der Identifier für ein Datum unabhängig von dessen Inhalt ist und nur vom öffentlichen Schlüssel und dem docName abhängt, könnte ein Subspace-Verwalter versuchen unter gleichem Namen eine neue Datenversion hochzuladen. Diese würde aber trotz gültiger Signatur von den anderen Nodes abgewiesen, da diese bereits ein Datum unter dem angegebenen Identifier besitzen.

USKs besitzen deshalb am Ende des docName ein zusätzliches Pfadelement, das eine Versionsnummer darstellt. Diese Versionsnummer kann nun entweder eine positive oder negative Ganzzahl sein. Hiervon ist der Suchalgorithmus für neue Versionen des Datums abhängig. Ist die Version positiv, sucht der Node in einer lokalen Datenbank, der *USK registry* nach dieser Version. Findet er dort eine neuere Version als angefragt wurde, liefert er diese aus. Gleichzeitig startet der Node eine Hintergrundsuche im *Freenet* nach neueren Versionen. Diese Suche funktioniert ebenso wie die Suche, die ausgeführt wird, wenn eine negative Versionsnummer angegeben wurde, mit der Ausnahme, dass das Aufrufen einer negativen Versionsnummer im Vordergrund passiert und dem Benutzer somit nicht sofort eine Version der von ihm angefragten Daten geliefert wird.

Für den Aktualisierungsalgorithmus versucht der Node die angefragte, sowie die vier nächsthöheren Versionen zu finden. Ist die Suche nach einer der fünf Versionen erfolgreich, sucht der Node nach fünf neueren Versionen. Dies geschieht so lange, bis er vier aufeinanderfolgende Versionen nicht finden kann. Ist zum Beispiel nach der Version -10 gefragt, sucht der Node nach den Version 10 bis 14. Findet er eine davon – unabhängig welche – fährt er mit der Suche nach 15 bis 19 fort. Wird dem Node nun explizit vom *Freenet* mitgeteilt, dass die Versionen 15 bis 18 nicht existieren, beendet er die Suche und liefert dem Benutzer die nun aktuellste Version aus der USK registry.

Auch beim Hochladen von Daten unter einer USK kann sich der Benutzer bei der Versionierung unterstützen lassen. Gibt man als Versionsnummer 0 an, sucht der Node die niedrigste, freie Nummer und benutzt diese beim Hochladen. Gibt man eine Version größer 0 an, versucht der Node erst das Datum unter dieser Version einzufügen. Ist sie bereits besetzt, wird automatisch nach der nächsten Freien gesucht und unter dieser eingefügt.

3.2 Einsatz von Kryptographie

Alle Daten, die in das *Freenet* hochgeladen werden, sind mit einem zufälligen, symmetrischen Schlüssel, dem cryptoKey, verschlüsselt. Da dieser aber nicht zusammen mit den Daten auf einem Node gespeichert wird, kann ein Node plausibel widerlegen, dass er Wissen über den Inhalt der von ihm gespeicherten Daten hatte. Inwiefern diese Unwissenheit vor Strafe schützt, kann wohl je nach Rechtssystem stark unterschiedlich angesehen werden. Der cryptoKey muss auf einem anderen Weg zusammen mit dem Identifier zugänglich gemacht werden, damit die Daten entschlüsselt werden können. Dazu kann beispielsweise die URI einer Datei ausgetauscht werden, die alle nötigen Informationen enthält.

Weitere Schutzmaßnahmen bestehen darin, dass auch die Kommunikation zwischen zwei Nodes verschlüsselt wird.

Es bestehen zudem Überlegungen, steganographische Transports zu implementieren. Dabei wird *Freenets* Datenverkehr in den Verkehr einer anderen, unauffälligen Anwendung eingebettet oder als solche "verkleidet". So soll verhindert werden, dass *Freenet* -Verkehr als solcher erkannt wird und damit blockiert werden kann.

Als Verschlüsselungsalgorithmus kommt AES mit einer Blockgröße von 256 Bit, anstelle der sonst üblichen 128 Bit zum Einsatz. Es gibt jedoch Überlegungen ebenfalls auf eine Blockgröße von 128 Bit umzusteigen, da diese Version von AES besser untersucht ist.

4. PROTOKOLLALGORITHMEN

4.1 Joining & Leaving

Der Beitritt in ein Peer-to-Peer-Netzwerk stellt immer eine sehr große Hürde da, insbesondere, wenn es wie bei *Freenet* keine zentralen Vermittlungsstellen geben soll. Gerade bei einem Dienst, der seinen Benutzern Anonymität anbieten möchte, muss verhindert werden, dass sich Nodes mit böswilligen Angreifern verbinden. Im Darknet-Modus ist dies dadurch sichergestellt, dass sich nur zu beidseitig manuell hinzugefügten Nodes verbunden wird, von denen man sich sicher ist, dass diese nicht böswillig sind.

Im Opennet-Modus, also wenn der Benutzer keine oder nur wenige Nodebetreiber kennt, denen er vertraut, ist dies ein erhebliches Problem. Um dennoch am *Freenet* teilnehmen zu können, wird eine Liste von sogenannten *Seed Nodes* angeboten. Die Liste kann beim Start eines Nodes über die *Freenet* -Webseite bezogen werden. Verbindet man sich mit einem Seed Node, vermittelt dieser anderen Nodes, die sich in der Umgebung der eigenen Location befinden, die nötigen Verbindungsdaten. Nodes, die an neuen Verbindungen interessiert sind, senden dann Verbindungsanfragen an den neu beitretenden Node.

Zur Optimierung der Netzwerktopologie und des Routings werden in Abhängigkeit des Modus, in dem der Node läuft, verschiedene Techniken angewandt. Im Darknet-Modus wird ein *Location Swapping*-Algorithmus angewendet, der auf [6] basiert. Dabei vergleichen zwei Nodes zu zufällig gewählten Zeitpunkten die mittlere Distanz zu den Nachbarn, mit denen sie jeweils verbunden sind. Wäre die mittlere Distanz geringer, wenn sie ihre Locations vertauschen würden, wird dieser Tausch vorgenommen. Dabei ändert sich nicht die Netzwerktopologie, also die Verbindungsstruktur zu anderen Nodes, da diese ja statisch konfiguriert wurde. Location Swapping kommt bei neuen Nodes noch relativ häufig vor;

die Frequenz lässt aber mit der Zeit nach, sofern sich die Verbindungen nicht ändern.

Da im Opennet eine dynamische Netzwerktopologie möglich ist, werden die Locations von Nodes nicht getauscht, sondern lebenslang beibehalten. Stattdessen werden wenig genutzte Verbindungen im Laufe der Zeit durch neue ersetzt.

Um aus dem Netzwerk auszutreten, sind keine speziellen Schritte notwendig. Da *Freenet* darauf ausgelegt ist widerstandsfähig gegenüber Soft- und Hardwareausfällen zu sein, verträgt es plötzlich verschwindende Nodes ohne Probleme.

4.2 Speichern & Abrufen von Daten

Als eine grundlegende Maßnahme gegen Profiling und Traffic Analysis[2], wurde die Datenblockgröße von CHKs auf 32 kB sowie von SSKs und damit auch USKs auf 1 kB beschränkt. Dies hat zudem den Vorteil, dass große Datenmengen parallel angefragt werden können und durch die zufällige Natur der Hashes auf viele Nodes verteilt werden.

Will ein Node nun doch eine Datei in das *Freenet* einbringen, welche größer als 32 kB ist, muss er diese aufteilen und das letzte Element gegebenenfalls auch auf 32 kB padden. Nun werden die einzelnen Teilstücke unabhängig voneinander hochgeladen. Zudem werden redundante Daten mit Hilfe von Vandermode Forward Error Correction Codes[4] eingefügt, um aufgrund der fehlenden Lebenszeitgarantie verlorene Teilstücke rekonstruieren zu können. Dabei bekommt jedes Teildatum einen eigenen CHK. Sind alle Teile im *Freenet*, erstellt der Node eine weitere Datei, die alle CHKs der eingefügten Teildaten beinhaltet und fügt diese zuletzt ein. So benötigen Benutzer lediglich diesen indirekten Dateizeiger, um die ursprüngliche Datei aufzurufen. Der *Freenet* - Node löst diese Indirektion für den Benutzer automatisch auf. Werden für den Dateizeiger USKs verwendet, kann zudem eine automatische Versionierung von Daten vorgenommen werden.

Um ein Datum im *Freenet* aufzufinden, wird eine *Greedy Search* nach dessen Identifier ausgeführt (bei SSKs und USKs Hash aus XOR-verknüpften routingKey und Hash des docName, bei CHKs der routingKey). Die Vorgehensweise ist hierbei wie folgt:

1. Der Node prüft, ob sich das Datum zum routingKey bereits in seinem lokalem Speicher befindet. Falls ja, ist die Suche trivialerweise beendet. Falls nicht, fährt er mit Schritt 2 fort.
2. Der Node schickt eine Suchanfrage an einen ihm bekannten Node, bei dem die Distanz zwischen dessen Location und dem routingKey am geringsten ist. Hierzu wird der Identifier ebenfalls als eine Fließkommazahl aus dem Intervall zwischen 0 und 1 dargestellt.
3. Der Node, welcher die Anfrage bekommen hat, prüft nun in seinem Speicher nach, ob er das Datum mit entsprechendem Schlüssel kennt. Ist dies nicht der Fall, fährt er ebenfalls wie in Schritt 2 fort.

Damit dies nicht zu einer unendlich langen Kette von Anfragen führt, enthält jede Anfrage ein *Hop-To-Live*-Feld (HTL).

Um den genauen Ursprung einer Anfrage zu verschleiern, wird der HTL nur mit einer bestimmten Wahrscheinlichkeit dekrementiert. Besitzt auch der Node die angefragten Daten nicht, bei dem der HTL auf 0 fällt, meldet er dies an seinen Vorgänger zurück. Der Vorgänger setzt dann mit dem Node fort, der die nächstgeringste Distanz zum routingKey hat. Sollte es zu einem Kreis im Routingpfad kommen, kann dies erkannt werden, da der angefragte Node selbst noch auf eine Antwort für den entsprechenden Identifier wartet.

Dies wird solange fortgesetzt, bis die Daten entweder gefunden wurden oder das gesamte Netzwerk mit einer Tiefe, die etwa der HTL entspricht[5], durchsucht wurde. Wurden die Daten gefunden, nehmen diese den umgekehrten Weg, der für die Anfrage zustande gekommen ist. Auf dem Weg zum Anfragensteller entscheidet jeder Node, ob er die Daten speichert, anhand dessen, ob er noch freien Speicherplatz hat. Ist dies der Fall, speichert er das Datum. Ansonsten, falls kein freier Platz mehr vorhanden ist, prüft er, ob der Eintrag, dessen Aufruf am längsten zurückliegt, eine größere Distanz zu seiner Location hat. Ist dies der Fall, wird der alte Eintrag durch den neuen ersetzt. Abbildung 2 zeigt dabei beispielhaft einen möglichen Suchablauf.

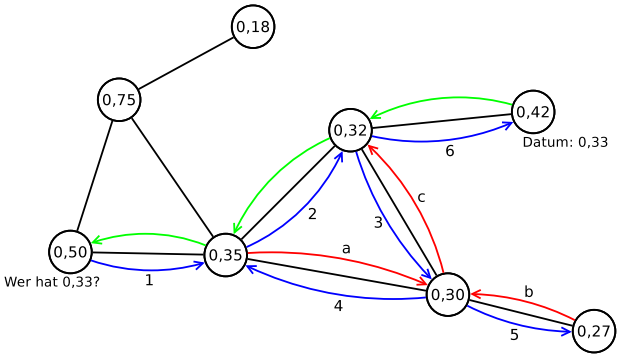


Abbildung 2: Ein möglicher Suchablauf. Der Node mit der Location 0,50 sucht nach einer Datei mit Schlüssel 0,33. Die blauen Pfeile geben die Folge der Anfragen an. Die roten sind Fehlermeldungen: *a*) Routingloop *b*) ist eine Sackgasse, Node 0,27 kann nicht weitersuchen *c*) der Node hat alle seine Nachbarn erfolglos befragt. Die grünen Pfeile zeigen den Rückweg der bei 0,42 erfolgreich gefundenen Daten. Die überquerten Nodes können sich entscheiden die Daten zu speichern.

Durch das Duplizieren der Daten auf dem Rückweg und des Überschreiben von unbenutzten Daten beziehungsweise Daten mit größerer Entfernung zur eigenen Location, werden mehrere Ziele erreicht. Einerseits verbreiten und verteilen sich „interessante", also oft aufgerufene Daten, im *Freenet*. Dies steigert die Zensurresistenz des Systems, da es beispielsweise nicht mehr nur einen zentralen Server gibt, auf dem die Daten gespeichert sind. Außerdem "spezialisieren" sich Nodes im Laufe der Zeit auf Daten, deren Identifier nahe ihrer Location liegen. Dadurch wird es wahrscheinlicher, dass Daten schneller, also über wenigere Hops gefunden werden können.

[5] Aufgrund der probabilistischen Dekrementierung wird tiefer als HTL gesucht

doi: 10.2313/NET-2015-03-1_02

Zum Hochladen von Daten startet ein Node genau so als würde er nach den Daten mit dem Schlüssel, den er einfügen möchte, suchen. Der Unterschied liegt darin, dass der letzte Hop nicht etwa eine Fehlermeldung zurückliefert, dass der Schlüssel nicht gefunden wurde, sondern ein O.K. gibt, dass der Schlüssel noch nicht existiert und der Einfügende beginnen kann die Daten zu senden. Auch hier speichern die Nodes auf dem Weg zwischen erstem und letzten Hop die Daten.

5. NUTZBARKEIT UND ANWENDUNGEN

Dem *Freenet* kann jeder durch das Betreiben eines *Nodes* beitreten. Der Installer lässt sich von *Freenets* Internetauftritt beziehen. Dabei handelt es sich um einen in Java geschriebenen Daemon, den man im Hintergrund laufen lassen kann. Auf das *Freenet* kann man mit einem handelsüblichen Browser zugreifen, da der Daemon einen Proxy, der standardmäßig auf Port 8888 des lokalen Rechners läuft, startet. Durch den Betrieb stellt man dem Netzwerk eine frei wählbare Speicherkapazität und Bandbreite zur Verfügung. Auf dem Konzept von *Freenet*, statische Daten zu speichern und abzurufen, wurden einige zusätzliche Dienste implementiert, die andere bereits bekannte Diensttypen bereitstellen.

Eine für andere Dienste grundlegende Anwendung ist das **Web of Trust** (WoT). Dieses hat keine Verbindung zum *Web of Trust*, das im Zusammenhang mit *OpenPGP* bekannt ist, welches aber wohl als Inspiration gedient haben wird. Es wird benutzt, um Vertrauen zwischen Identitäten oder Pseudonymen im *Freenet* herzustellen. Das WoT ermöglicht es, beliebige Identitäten zu erstellen, repräsentiert durch ein asymmetrisches Schlüsselpaar. Um andere Identitäten kennen zu lernen, veröffentlicht jeder Benutzer eine Liste von ihm bekannten Pseudonymen. Zusätzlich kann jedem Pseudonym ein Trustlevel zugeteilt werden, das angibt, für wie vertrauenswürdig man eine Identität hält. Das WoT wird in den anderen Diensten vor allem zur Erkennung von Spammern benutzt.

Weiter gibt es einen Email-Service, genant **Freemail**, der es erlaubt vertrauliche Email zu verschicken.

Frost und das **Freenet Message System** (FMS) sind zwei newsgroup-ähnliche Dienste, wobei Frost das ältere System ist und FMS zum Ziel hatte, viele Probleme zu beheben die Frost plagen. So hat Frost kein System um Spammer zu erkennen und zu blockieren. Im FMS funktioniert dies ähnlich wie im WoT über das Veröffentlichen von Trust- Listen. Unterschreitet ein Pseudonym ein bestimmtes Trustlevel, werden dessen Nachrichten ignoriert und nicht vom Node abgefragt.

Sone ist ein Facebook oder Twitter ähnelndes Social Network, auf dem kurze Beiträge gepostet werden. Es benutzt ebenfalls WoT zum Auffinden und Verwalten von Identitäten.

6. PROBLEME UND AUSBLICK

Ein großes Problem ist die Suche nach für den jeweiligen Nutzer relevanten Inhalten und das Auffinden der zum Entschlüsseln verwendeten cryptoKeys. So existiert etwa ein Crawler, welcher versucht *Freenet* -Seiten und darauf enthaltene Keys zu indizieren um diesen Index dann durchsuchen zu können. Zudem existieren mehrere manuell gepflegte Seitenindizes, welche regelmäßig aktualisiert werden.

Leider kann das Hoch- und Herunterladen von größeren Datenmengen mehrere Stunden oder Tage in Anspruch nehmen. Einfache Webseiten mit nur wenigen Bildern laden in der Regel innerhalb von einigen 10 Sekunden. Außerdem gibt es aufgrund der Architektur keine Garantie, dass Daten, die man heute einfügt, auch einige Zeit später noch abrufbar sind.

Eine Evaluierung des Routingalgorithmus findet in [3] statt. Es wird zudem eine Attacke auf das Location Swapping demonstriert, die es einem beliebigen Angreifer erlaubt, das Netzwerk so stark zu stören, dass es zu signifikantem Datenverlust kommen kann.

Roos et al. haben erfolgreich versucht das bestehende *Freenet* zu vermessen[5]. Damit wurde gezeigt, das einige Obfuscationmaßnahmen in *Freenet* nicht halten, was sie versprechen, da sie solche Untersuchungen verhindern hätten sollen.

Damit *Freenet* wirklich effizient und mit größtem Vertrauen in dessen Netzwerk benutzt werden kann, ist es vonnöten, das Konzept des Darknets weiter auszubauen und zu verbreiten. Wie sich dieses Bootstrapping-Problem für Menschen, die beispielsweise von staatlicher Seite verfolgt werden und deshalb auf einen sicheren und anonymen Kommunikationskanal angewiesen sind, beheben lässt, ist noch offen. Dabei kann auch schon das vertrauensvolle Beziehen der Software problematisch sein, sollte der Angreifer in der Lage sein, Man-in-the-Middle-Attacken auszuführen.

Zudem muss sich zeigen, wie gut *Freenet* in der realen Welt skalierbar ist, sollte es größere Adaption erfahren. Noch wird laut [5] *Freenet* vorallem in westlichen Ländern wie den USA, Deutschland oder Großbritannien verwendet, die noch nicht so stark von staatlicher Zensur betroffen sind. In Ländern, deren Regierungen öfter Zensur und staatliche Unterdrückung vorgeworfen wird, ist *Freenet* noch nicht sehr populär.

7. VERWANDTE ARBEITEN

Freenet lässt sich nur eingeschränkt mit Anonymisierungsdiensten wie *TOR* vergleichen. Im Gegensatz zu TOR gibt es keine Möglichkeit Daten von außerhalb des Netzwerkes aufzurufen. TOR bietet über sogenannte *Exit Nodes* die Möglichkeit das “normale”, öffentliche Internet aufzurufen. Zusätzliche Webseiten, die *Hidden Services* sind nur innerhalb von TOR routbar. Im Gegensatz zu *Freenet*, sind dies aber keine verteilten Datenspeicher, sondern nur speziell konfigurierte Server. Dies macht sie, sobald man deren echte IP-Adresse ermittelt hat, anfällig für Zensur und DDoS-Attacken.

Ein weiteres anonymes Netzwerk ist das *Invisible Internet Project* (I2P). Wie TOR besitzt es *Hidden Services*, die nur aus dem I2P-Netzwerk erreichbar sind. Es findet aber auch hier keine Duplizierung von Daten auf verschiedenen Nodes statt. Im Gegensatz zu *Freenet* bietet aber auch I2P die Möglichkeit das normale Internet anonym zu benutzen. Somit ist seine Ähnlichkeit zu TOR bedeutend grüßer als zu

doi: 10.2313/NET-2015-03-1_02

Freenet.

Ein Projekt mit anfangs ähnlichen Zielen wie *Freenet* ist *GNUnet*. Es wurde als anonymes, zensurresistentes Filesharingprogramm entworfen, ist aber seither stark gewachsen und legt den Fokus nun mehr darauf ein allgemeines, dezentrales Netzwerk aufzubauen. So bietet *GNUnet* ein alternatives DNS, genannt *GNU Name System*, sowie eine dezentrale Public Key Infrastruktur (PKI).

8. ZUSAMMENFASSUNG

In dieser Arbeit wurde ein Überblick über *Freenet* gegeben. Es handelt sich hierbei um ein anonymes Peer-to-Peer-Netzwerk, das zum Ziel hat, eine verteilte, zensurresistente Möglichkeit zum Speichern und Abrufen von Daten im Internet zu bieten.

Die Zensurresistenz beruht darauf, dass Daten ständig, auf viele Nodes verteilt, dupliziert werden. Dadurch ist *Freenet* resistent gegen gezielte Attacken auf einzelne Nodes. Dies hat aber auch zur Folge, dass es keine Garantie gibt, dass wenig genutzte Daten im *Freenet* für längere Zeit aufrufbar bleiben.

Für das Routing im Darknet-Modus, welcher der bevorzugte Modus gegenüber dem Opennet ist, wird argumentiert, dass sich ein sognanntes *Small World Network* bildet. Der Grund hierfür ist, dass sich die sozialen Bekanntschaftsstrukturen im *Freenet* widerspiegeln, wenn nur Verbindungen mit vertrauenswürden Freunden hergestellt werden. Dies wurde bei der Implementierung des Routingalgorithmus beachtet, weshalb dieser insbesondere in einem *Small World Network* effizient arbeiten kann.

Zusätzlich zum *Freenet* existieren teils von diesem unabhängige Projekte, um zusätzliche Dienste wie Email oder Newsgroups anzubieten.

9. LITERATUR

[1] I. Clarke and S. D. C. Mellish. A distributed decentralised information storage and retrieval system. Technical report, 1999. `https://freenetproject.org/papers.html`.

[2] I. Clarke, O. Sandberg, M. Toseland, and V. Verendel. Private communication through a network of trusted connections: The dark freenet. `https://freenetproject.org/papers.html`.

[3] N. S. Evans, C. Gauthierdickey, and C. Grothoff. Routing in the dark: Pitch black. In *In Proceedings of the 23rd Annual Computer Security Applications Conference (ACSAC). IEEE Computer Society*, 2007.

[4] L. Rizzo. Effective erasure codes for reliable computer communication protocols, 1997.

[5] S. Roos, B. Schiller, S. Hacker, and T. Strufe. Measuring freenet in the wild: Censorship-resilience under observation. In E. De Cristofaro and S. Murdoch, editors, *Privacy Enhancing Technologies*, volume 8555 of *Lecture Notes in Computer Science*, pages 263–282. Springer International Publishing, 2014.

[6] O. Sandberg. Distributed routing in small-world networks, 2005. `https://freenetproject.org/papers.html`.

Timing of Cyber Conflict

Fabian Heidler
Betreuer: Heiko Niedermayer
Seminar Future Internet WS2014
Lehrstuhl Netzarchitekturen und Netzdienste
Fakultät für Informatik, Technische Universität München
Email: ga46zuy@in.tum.de

Diese Arbeit soll darüber Einblick geben wann der richtige Zeitpunkt gekommen ist, eine Lücke in einem fremden Sicherheitssystem auszunutzen. Dazu wird ein grundlegendes mathematisches Modell erklärt und anschließend an einigen Beispielen demonstriert. Des Weiteren soll die Bedeutung von Cyber Conflict erkannt werden, und Ausblick gegeben werden welche Rolle dieser in der Zukunft einnimmt. Abschließend wird noch eine mögliche Anwendung in Computerspielen diskutiert.

1. Einleitung

1.1 Was ist Cyber Conflict

Cyber Conflict ist das Nutzen von Computer Ressourcen um Informationen zu beschaffen oder Schaden zu verursachen. Die Ressourcen die Staaten oder Organisationen hier ansammeln werden Exploits genannt. Dies ist eine Möglichkeit eine Schwachstelle in einem System auszunutzen, die bei der Implementierung übersehen wurde. Dies ist nicht zu verwechseln mit der Vulnerabilität, welche die Sicherheitslücke darstellt. Exploits werden sowohl zum Diebstahl von Daten genutzt, aber auch zur simplen Sabotage an Hardware . Ein Zero – Day – Exploit beschreibt einen möglichen Angriff auf eine Schwachstelle, der bereits sehr kurz nach Release der Software entwickelt wurde und folglich in dieser Form noch nicht bekannt ist. Wird diese Schwachstelle nicht direkt von den Entwicklern selbst entdeckt, besteht die Möglichkeit dass diese lange unentdeckt bleibt. Das macht Zero – Day – Exploits so wertvoll. Der Direktor der National Security Amerikas, nannte Cyber Security als wichtigste Bedrohung der sich Amerika gegenüber sieht. Nachdem immer mehr Systeme sich ausschließlich auf Software verlassen, wird also auch die Bedeutung von Cyberkrieg wichtiger. Der Schaden der hierbei entstehen kann lässt sich in 6 Felder einteilen:

- Verlust von geistigem Wissen und vertraulichen Informationen
- Cyberkriminalität,
- Diebstahl von vertraulichen Geschäftsinformationen und damit die eventuelle Manipulation von Börsendaten
- Kosten für die Verteidigung der Netzwerke sowie Versicherungskosten
- Rufmord an betroffenen Firmen[1]
- Opportunitätskosten

Anzumerken ist noch, dass ein wirksamer Angriff häufig nicht nur aus einem Exploit besteht. Eine sinnvolle Ressource, kann also mehrere Exploits beinhalten und greift eventuell auch auf nicht technische Hilfsmittel zurück, zum Beispiel einen Insider, der den Angriff einschleust.

1.2 Ablauf eines Exploits

Den Anfang macht immer das unabsichtliche Einschleusen eines Bugs in das Programm. Das Programm wird also mit Schwachstellen an die Verbraucher gebracht. Irgendwann wird die Schwachstelle entdeckt, und ein Exploit wird entwickelt um diese auszunutzen, meistens von Kräften aus der Unterwelt. Nach dem - für eine gewissen Zeitspanne - heimlichem Einsatz, also der Zero – Day - Attack, erfährt der Entwickler von der Lücke, entweder durch Tests oder durch Meldungen von Nutzern, und beginnt an einem Patch zur Behebung zu arbeiten. Kurz darauf wird die Vulnerabilität an die Öffentlichkeit weitergegeben, was zu Follow – on – attacks führt. Dadurch wissen nun auch Anti-Virus Hersteller von der Lücke und updaten ihre Programme um diese zu erkennen. Endnutzer mit aktuellem Anti-Viren Programm können nun erkennen, ob sie infiziert sind. Kurz nach Veröffentlichung durch den Hersteller, wird der Patch freigegeben und die Nutzer der aktuellen Version sind wieder geschützt (siehe auch Bild 1).

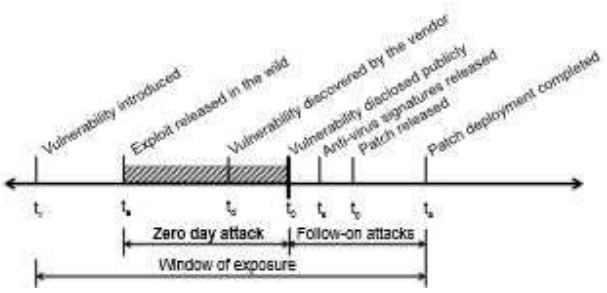


Bild 1. Zeitlicher Ablauf eines Zero – Day – Exploits [2]

In dieser Arbeit soll nun ein Model von Robert Axelrod und Rumen Iliev näher gebracht werden, welches sich damit auseinandersetzt, wann der Zeitpunkt gekommen ist, einen Zero – Day – Exploit einzusetzen. Das Schwierige dabei ist abzuschätzen, ob es sich bereits lohnt seinen Exploit zu zeigen, wobei dieser dann möglicherweise nicht mehr einsetzbar ist. Auf der anderen Seite kann zu langes Warten dazu führen dass die Schwachstelle entdeckt wird und der Exploit nutzlos geworden ist. Da dieses Model sehr mathematisch ist, werden zum besseren Verständnis drei Ereignisse gezeigt: der iranische Angriff auf Saudi Aramco, die tägliche Cyber Spionage Chinas und ein frühzeitiger Einsatz Chinas gegen Japan.

doi: 10.2313/NET-2015-03-1_03

2. Mathematisches Modell

2.1 Erklärung

2.1.1 Der Einsatz

Die erste Variable die zu berücksichtigen ist, stellt der Einsatz dar. Oder in anderen Worten: Wie viel steht zum derzeitigen Zeitpunkt auf dem Spiel. Das Problem hierbei ist, dass zwar bekannt ist was im Moment der Einsatz ist, jedoch lässt sich keine Aussage darüber treffen wie sich dieser entwickelt. In Kriegszeiten hat eine Ressource also einen deutlich höheren Wert, als wenn Frieden herrscht. Doch auch ohne Konflikt kann ein Land Interesse haben an der Technologie eines anderen. Dann würde der Einsatz auf einem mittleren Level stehen. Insgesamt ist der Einsatz durch die gesamte Vernetzung der Welt und unvorhersehbaren politischen Ereignissen am schwersten einzuschätzen.

2.1.2 Stealth und Persistence

Das Überleben einer Ressource hängt maßgeblich von 2 Faktoren ab. Zum einen von ihrer Tarnung („Stealth) und zum anderen von ihrer Beständigkeit („Persistence“). Stealth gibt dabei an, wie hoch die Wahrscheinlichkeit ist, dass nach Benutzung der Ressource, sie unentdeckt bleibt und somit wieder einsetzbar ist. Ein Beispiel für die Tarnung ist der Conficker Wurm, der bis zu seiner ersten Entdeckung im Oktober 2008 über 370,000 Computer infizierte. Die geschätzte Nummer an Betroffenen bis zum Januar 2009 reicht von 9 Millionen bis 15 Millionen. Damit ist er der größte Wurm der zur Zeit bekannt ist, seit dem Welchia Wurm 2003. Inzwischen gibt es aber genügend Werkzeuge für seine Entfernung.[3]

Persistence hingegen, ist die Wahrscheinlichkeit dafür, dass eine unbenutzte Ressource auch weiterhin unentdeckt bleibt. Die Vulnerabilität wird also nicht behoben, und der Exploit kann weiterhin benutzt werden. Natürlich ist es auch schwer für diese beiden Variablen feste Werte zu finden, allerdings kann man anhand von vorangegangen Attacken und Beispielen ungefähre Schätzwerte festlegen. So liegt die Durchschnittliche Haltbarkeit von Zero – Day – Lücken bei 312 Tagen. Allerdings kann diese auch wesentlich länger sein. So wurden bei den Browsern Chrome und Firefox in 3 Jahren, nur eine geringe Anzahl an Lücken unabhängig entdeckt. Was die Persistence nahezu auf 1 heben würde.

Natürlich hängen diese beiden Variablen nicht nur von der Qualität des Exploits ab, sonder auch davon, wie wachsam und gut geschützt das gewählte Ziel ist. Gegen jemanden der seine Sicherheitspatches stets up – to – date hält ist die Persistence geringer, als gegen ein Ziel das seine Sicherheit vernachlässigt.Gleichwohl ist die Stealth höher gegen ein Ziel, welches wenig Wachsamkeit zeigt.

2.1.3 Der richtige Zeitpunkt

Die letzte Variable die noch zu berücksichtigen ist, stellt die Discount Rate *w* dar, in unserem Sinne die Inflation. Der Wert einer Information ist also im nächsten Jahr, nicht mehr genau so interessant, wie er es zum jetzigen Zeitpunkt ist. Die Rate liegt also immer im Bereich zwischen 0 und 1.

Das einzige was nicht unter der Kontrolle des Angreifers liegt ist der Einsatz. Deswegen ist es sinnvoll so lange mit den Nutzen der Ressource zu warten, bis dieser hoch genug ist dass man sie aufgeben kann. Also legt man eine Grenze fest, ab deren Überschreiten es sich lohnt zu zuschlagen. Dabei ist der Gewinn *G(T)*, den man bei linearem Einsatz zieht, nur bei einem Grenze von *a* oder *b* anzugreifen: *G(T) = (a + b) / 2*. Je geringer man also die Grenze setzt, desto öfter kann man seine Ressource zwar einsetzen, aber dadurch bleibt der durchschnittliche Gewinn ebenfalls klein. Das grundlegende Problem ist, die Ressource möglichst oft einzusetzen, aber sie gleichzeitig für Zeiten sparen, in denen viel auf dem Spiel steht.

Wird die Ressource zum jetzigen Zeitpunkt eingesetzt, ergibt sich ihr Wert *V* aus dem erwarteten Gewinn dieses Nutzen und dem Zukunftswert, der von der Stealth *S* sowie der Discount Rate abhängt. Damit folgt für die akute Nutzung: *V(Nutzung der Ressource) = G(T) + w S V* .Hebt man sich die Ressource auf, so errechnet sich der Wert aus der Wahrscheinlichkeit, dass sie auch weiterhin Bestand hat, und ebenfalls der Disount Rate. Folglich ist hier die Gleichung: *V (Aufsparung der Ressource) = w P V* .Die Chance dass eine Ressource eingesetzt wird, ist die Wahrscheinlichkeit *Pr* dafür, dass der aktuelle Einsatz mindestens so groß ist wie die festgelegte Grenze, *Pr(s >= T).* Analog dazu ist die Chance, dass sie aufgespart wird, die Gegenwahrscheinlichkeit, *1 - Pr(s >= T)*

Setzt man die einzelnen Bestandteile zusammen, erhalten wir unseren erwarteten Wert der Ressource:

$$V = Pr(s \geq T)(G(T) + w\,S\,V) + (1 - Pr(s \geq T))\,w\,P\,V$$

Natürlich ist es sinnvoller den Wert auf einer Seite gesondert zu haben, denn schließlich weiß der Angreifer nur Stealth, Persistence und eine selbstgewählte Verteilung des Einsatzes. Deswegen erhalten wir nach Umformen unserer ursprünglichen Gleichung.

$$V = \frac{(Pr(s \geq T)\,G(T))}{((1 - w * P) + Pr(s \geq T)\,w(P - S))}$$

Es ist nun möglich den Wert einer Ressource anhand dieser Formel zu berechnen. Aus diesem Wert lässt sich schließen, wann der beste Zeitpunkt gekommen ist anzugreifen.

2.2 Anwendung des Modells

Wir wenden nun die Rechnung zum besseren Verständnis an. Zuerst an einem einfachen Beispiel, in dem der Einsatz linear verteilt ist. Die Dringlichkeiten 1, 2, 3 ,4 ,5 ,6 des Einsatzes treten alle mit gleicher Wahrscheinlichkeit auf. Außerdem sei die Disount Rate auf 0.8 festgelegt und die Stealth auf 40% der Persistence. Wendet man nun die Formel an, sieht man wie sich die Persistence auf den optimalen Grenzwert auswirkt. Die verschiedenen Ergebnisse sind in Tabelle 1 dargestellt. Um die Rechnungen nachzuvollziehen, wird im Folgenden gezeigt wie wir auf den Wert der Tabelle im Feld 0,1 – 1 kommen. Unser *G(T)* ist in diesem Fall *(1+2+3+4+5+6) / 6 = 3,5*. Da unser Grenzwert eins ist, nutzen wir die Ressource bei jeder Gelegenheit also ist *Pr (s>= T) = 1*. Die Restlichen Werte sind bekannt, also ist es nun simples Einsetzen.

$$V = \frac{(1 * 3{,}5)}{((1 - 0{,}8 * 0{,}1) + 1 * 0{,}8(0{,}1 - (0{,}4 * 0{,}1)))} = 3.616$$

doi: 10.2313/NET-2015-03-1_03

T	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1
6	1,08	1,17	1,28	1,4	1,56	1,76	2,02	2,36	2,84	3,57
5	1,96	2,1	2,27	2,46	2,7	2,98	3,32	2,76	4,32	5,09
4	2,65	2,81	3	3,22	3,47	3,77	4,11	4,53	5,04	5,68
3	3,15	3,32	3,5	3,71	3,95	4,21	4,52	4,87	**5,28**	**5,77**
2	3,47	3,62	3,79	3,97	**4,17**	**4,39**	**4,63**	**4,9**	5,2	5,56
1	**3,62**	**3,74**	**3,87**	**4,01**	4,17	4,33	4,51	4,7	4,92	5,15

Tabelle 1. Die Auswirkungen der Persistence

Die Spalten stehen für die derzeitige Persistence und die Zeilen für den Grenzwert

Aus den Ergebnissen lässt sich herauslesen, dass je höher die Persistence ist, desto länger es sich lohnt mit dem Angriff zu warten. Hat man jedoch eine geringe Persistence empfiehlt es sich die Ressource sofort zu nutzen. Logisch ist das ebenfalls verständlich, denn die Wahrscheinlichkeit dass die Ressource verfällt, auch bei Nichtnutzung, ist sehr hoch. Auch bei anders gewählten Zahlenbeispielen stieg der optimale Grenzwert, als die Persistence sich vergrößerte. Die nächste Tabelle zeigt den Effekt den die Stealth auf den optimalen Grenzwert hat. Dabei gehen wir von einer konstanten Persistence von 0,8 aus, die Discount Rate belassen wir auf 0,8.

T	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1
6	2,21	2,72	2,34	2,42	2,5	2,59	2,68	2,78	2,88	3
5	3,35	3,53	3,72	3,93	4,17	4,44	4,74	5,09	5,5	5,98
4	3,91	4,17	4,46	4,81	5,21	5,68	6,25	6,94	7,81	8,93
3	**4,09**	**4,41**	**4,89**	5,23	5,77	6,43	7,26	8,33	9,78	11,8
2	4,03	4,39	4,81	**5,32**	**5,95**	**6,76**	7,81	9,26	11,4	14,7
1	3,8	4,17	4,61	5,15	5,83	6,73	**7,95**	**9,72**	**12,5**	**17,5**

Tabelle 2. Die Auswirkungen von Stealth

Die Spalten stehen für den derzeitigen Stealth und die Zeilen für den Grenzwert

Stealth zeigt genau den gegensätzlichen Effekt. Je höher der Stealth der Ressource ist, desto mehr lohnt es sich diese möglichst früh einzusetzen. Der Sinn dahinter ist, dass der Angriff aufgrund seiner hohen Tarnung nicht entdeckt wird, und somit wesentlich öfter ausgeführt werden kann. Das soll allerdings nicht bedeuten, dass es bei hohem Einsatz sinnvoller ist nicht auf den Stealth zu achten. Sowohl Stealth als auch Persistence sind wertvolle Eigenschaften, die ein Exploit besitzen kann. Aber daraus lässt sich schließen, dass eine Ressource, die nur eine sehr geringe Tarnung hat, besser für Zeiten aufgespart werden sollte, in denen viel auf dem Spiel steht, da sie mit großer Wahrscheinlichkeit nur einmalig einsetzbar ist.

Natürlich gibt es noch den wesentlich einfacheren Fall wenn man einfach mit konstantem Einsatz rechnet. Das Risiko bleibt also immer gleich hoch, dass trifft zum Beispiel auf Kriminelle zu, deren Gewinn aus dem Hacken von Kreditkarten besteht. Genauso verhält es sich mit terroristischen Organisationen. Besteht deren Ziel einzig allein darin möglichst viel Schaden zu verursachen, werden sie nicht auf besondere Ereignisse warte, sondern so oft Anschläge verüben, wie es ihnen möglich ist. Ist der Einsatz also gleichbleibend, ist die beste Taktik seine Ressource so oft und so lang wie es geht einzusetzen.

Doch meistens sind die Einsätze wohl eher ungleich verteilt und die wichtigen Ereignisse, sind zwar sehr selten, aber falls sie eintreffen, übertreffen sie die üblichen Zustände bei weitem. So könnten die Einsätze zum Beispiel, bei 1, 4, 9, 16, 25, 36 liegen ,in Form einer Parabelfunktion, oder aber auch als exponentielle Verteilung, 1, 2, 4, 8, 16, 32. In unserem Model hätte das Einfluss auf *G(T)* und *Pr(s >= T)*. Gehen wir einmal von einer exponentiellen Verteilung aus, so wäre *G(32) = 32*, aber die Wahrscheinlichkeit *Pr(s>=T)*, dass diese Grenze je überschritten wird, läge nur bei 1/32. Anhand des Modells, lässt sich sagen, dass je mehr die Einsätze verzerrt sind, desto mehr steigt der optimale Grenzwert. Aber desto länger muss man auch warten, bis dieser Wert überschritten wird. Gerade bei stark verteilten Einsätzen muss man allerdings darauf achten, nicht zu lange zu warten, denn dass der höchste Fall erreicht wird, ist sehr unwahrscheinlich.

T	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1
32	1,1	1,22	1,36	1,55	1,8	2,13	2,63	3,43	4,93	8,77
16	13,7	14,7	15,9	17,3	19	21,1	**23,7**	**27,7**	**31,5**	**37,6**
8	**15,5**	**16,4**	**17,5**	**18,7**	**20,1**	**21,8**	23,7	26	28,9	32,4
4	14,3	15,1	16	16,9	18	19,3	20,7	22,4	24,4	26,8
2	12,6	13,2	14	14,8	15,6	16,7	17,8	19,1	20,6	22,4
1	11	11,5	12,1	12,8	13,6	14,4	15,3	16,4	17,7	19,1

Tabelle 3. Verteilte Einsätze

Die Spalten stehen für die Persistence und die Zeilen für den Grenzwert, w = 0,9 S = P/2

Bevor nun die echten Fallbeispiele kommen, lässt sich zusammenfassen, dass es 3 Faktoren gibt, die eine Ressource auszeichnen, mit deren Einsatz es sich lohnt zu warten. Diese wären, eine niedrige Stealth, eine hohe Beständigkeit und die Chance einen großen Nutzen in Fällen, in denen viel auf dem Spiel steht zu ziehen. Umgekehrt ist eine Ressource mit hoher Tarnung, aber wenig Beständigkeit besser dafür geeignet, in alltäglichen Dingen eingesetzt zu werden.

3. Geschichtliche Fallbeispiele

3.1 Iranischer Angriff auf Saudi - Aramco

Bei diesem Angriff war die Saudi – Arabischen Ölfirma Saudi Aramco das Ziel. Diese ist die derzeit größte Erdölfördergesellschaft der Welt und fördert jährlich 525,0 Mio. Tonnen Erdöl. Die Financial Times nannte die Firma mit 10 Billion US – Dollar Unternehmenswert sogar das wertvollste Unternehmen der Welt.[4] Am 15ten August 2012 infizierte ein Computervirus die Firma, und befiel ungefähr 30.000 windowsbasierte Rechner. Später wurde dieser Shamoon getauft.

Seine Hauptfunktion bestand wohl darin Daten von Computerfestplatten zu löschen. Der Virus bestand dabei aus 3 Teilen, einem Dropper, der sich auf den Hauptteil bezog und Quelle der Infektion war, einem Wiper – Modul, das für die Zerstörung der Daten zuständig war und einem Reporter Modul, das die Infektion an den Angreifer zurückmeldete. Der Virus überschrieb dabei die Daten mit einem Bruchteil eines Bildes, das eine brennende amerikanische Flagge zeigte.[5] Obwohl der Virus erfolgreich 75% der Arbeitsstationen infizierte[6] beschädigte er nicht die kritische Infrastruktur, da diese auf isolierten System lagen. Es gab also keinen Ölverlust oder Explosionen, dennoch ist davon auszugehen dass Bohr und Produktionsdaten verloren gegangen sind.

Obwohl es einer der größten Angriffe auf eine einzelne Firma war, blieb also der große Schaden aus. Das lag vor allem daran dass die Attacke nicht sehr unauffällig vorging und schnell gestoppt wurde. Nach 4 Tagen war der Virus komplett entfernt. Zuerst bekannte sich eine Hackergruppe mit dem Namen „Cutting Sword of Justice“ zu dem Angriff, allerdings wurde nach genaueren Recherchen klar dass die Quelle aus dem Iran kam. Die Regierung lehnte zwar jegliche Verantwortung ab, aber die Kontrolle über das Internet im Iran ist so streng, dass es schwer vorstellbar ist, dass sie davon nichts gewusst hätte. Ebenfalls dafür spricht, dass der Iran kurz davor selber von einer Attacke getroffen wurde, dem Stuxnet, welche dem iranischen Atomprogramm schadete. Vermutlich fühlte sich der Iran dazu gedrängt schnell eine ebenbürtige Antwort zu senden , um nicht als schwach angesehen zu werden. Das erklärt auch einige Fehler im Virus. Die niedrige Stealthrate und die Tatsache, dass viel auf dem Spiel stand, decken sich ebenfalls mit unserem Modell, was unter solchen Bedingungen auch den sofortigen Einsatz einer Ressource vorschlägt.

Jedoch war der Angriff auch gleichzeitig ein Weckruf an die Firma ihre Sicherheit zu erhöhen. Saudi Aramco deckt ein Zehntel des Weltbedarfs an Öl.[7] Sollte deren Produktion stark geschädigt werden, wäre das ein internationaler Schaden. Das zeigt noch einmal was für eine Bedrohung Cyber Conflict darstellt.

3.2 Tägliche E-Spionage Chinas

China stellt eine außergewöhnlich Rolle im Cyberkrieg dar. Zwar besitzt keine Nation eine reine Weste und jegliche Verknüpfung mit E-Spionage wird verleugnet, was dazu geführt hat, dass diese Aussage nur noch als Fiktion angesehen wird, dennoch übertrifft China viele seiner Konkurrenten. Einigen Schätzungen zufolge haben 90% der Angriffe die in der USA erfolgen ihren Ursprung in China(8). Dabei scheint dass Hauptziel der Diebstahl von Technologie zu sein, aber gleichzeitig besteht auch die Gefahr dass die militärischen Kapazitäten eines Landes ans Licht geraten. Ein 2013 veröffentlichter Report des Pentagons besagt, dass die China inzwischen soviel in seine Cybertechnologien investiert hat, dass es nun eine führende Rolle einnimmt. Dabei belaufen sich die Ausgaben von China selbst für Abwehrstrategien auf 135 bis 215 Milliarden Dollar, dies ist aber sogar mit den höchsten Werten nur ein Drittel dessen was die USA investiert.[8] Aber auch China beschuldigt die USA Spionage zu betreiben, eine Anschuldigung, die auch durch die Leaks von Edward Snowden bekräftigt wurde. Am 5. Mai 2014 klagte die USA chinesische Offiziere wegen Wirtschaftsspionage an. Als Gegenantwort rief China den US- Außenbotschafter zu sich. Dies sind natürlich alles vorerst symbolische Akte, es wird weiterhin auf Verhandlungen gesetzt. Aber das Verhältnis zu China bleibt dadurch angespannt. Tom Denilon, der Sicherheitsberater nannte das Lösen dieser Probleme den Schlüssel für ein zukünftiges gutes Verhältnis zu China.[9] US – Justizminister Holder meinte, außerdem dass der wirtschaftliche Erfolg eines Landes nicht davon abhängen dürfe, wie gut deren Ressourcen zur Spionage genutzt werde.[10]

Ein Grund dafür, dass China so oft erwischt wird, bei ihren Versuchen an Informationen zu gelangen, liegt daran dass die Stealth oft nur sehr durchschnittlich gehalten wird. Das wird zum Einen, durch die Vielzahl an Attacken ausgeglichen, sodass manche Ziele überlastet sind, zum Anderen, ist nicht jedes Ziel auf dem neuesten Stand der Technik.

Betrachten wir die chinesischen Angriffen anhand unseres Modells, scheint der häufige Einsatz der Ressource nicht gerechtfertigt, da der Einsatz im Moment doch sehr gering ist. Inzwischen ist China kein Entwicklungsland mehr und stellt auch eine starke wirtschaftliche Macht dar, womit der technologische Informationsgewinn nicht mehr so hoch ist, wie er es vor 20 Jahren war. Ebenfalls gibt es keinen militärischen Konflikt mit anderen Ländern. Ein möglicher Grund warum sie nicht warten, ist, dass sie bei all ihren Ressourcen von geringer Überlebensdauer ausgehen, also niedriger Persistence. Eventuell besitzt China auch ausreichend bessere Ressourcen, sodass es ihnen nicht schadet, die schwächeren sofort und oft einzusetzen. Eine letzte Überlegung ist noch, dass sie sich gegen schlechter geschütztere Zielen eine höhere Stealth erwarten, und somit über mehrere Jahre Informationen sammeln können. Die hohe Stealth einer Ressource schlägt wie Tabelle 2 zeigt einen häufigen Einsatz dieser vor.

3.3 Frühzeitiger Einsatz einer Ressource am Beispiel China

In den vorangegangenen Beispielen, fand der Einsatz einer Ressource immer zum erwarteten Zeitpunkt statt, gemessen an unserem Modell. Doch es gibt auch Ereignisse in denen Länder zu früh gehandelt haben, dabei muss es sich bei der Ressource nicht immer um einen Cyberexploit handeln. So geschehen, bei dem Exportstopp Chinas von seltenen Erdelementen um wirtschaftlichen Druck auf Japan auszuüben.

Zu den Seltenen Erdelementen zählen insgesamt 17 Elemente. Deren Name rührt daher, dass große Lagerstätten, also Ansammlungen von diesen selten sind. Ein Großteil der Gewinnung, besteht daher aus der chemischen Aufbereitung bei Metallen, die häufiger in der Erden vorkommen. Folglich ist es sehr aufwändig Minen zu errichten und die seltenen Erdelemente zu gewinnen. Bis in die 1980er Jahre hinein, war die US führender Produzent, stellte dann allerdings immer mehr den Minenbetrieb ein, denn zu diesem Zeitpunkt wurden sie hauptsächlich für Forschung oder ganz spezielle Aufträge verwendet. Andere Länder folgten diesem Beispiel. Nur China widerstand der Versuchung und hielt die Produktion aufrecht, was damals wenig Sinn zu ergeben schien.[11] Heutzutage werden seltene Erdmetalle in vielen Schlüsseltechnologien eingesetzt. Sie finden ihre Anwendung in Solaranlagen, Computerbildschirmen, Legierungen und vielen weiteren Bereichen. Lange Zeit war der chinesische Export billiger als selbst wieder in Eigenproduktion zu treten. Im Jahr 2009 wurden 124.000 Tonnen von ihnen verwendet, mit jährlich steigender Nachfrage. China hat auf das richtige Pferd gesetzt und nimmt einen Anteil von über 90% an

den Exporten von seltenen Erdmetallen ein. Diese Monopolstellung erweist sich in unserem Fall als wichtiges Druckmittel.

Am 7. September 2010 kollidierte ein Fischerboot mit 2 Japanischen Küstenwache Schiffen im Seegebiet nahe den Senkaku Inseln.[12] China und Japan streiten schon länger um diese Inseln, deswegen nahmen die Japaner die chinesische Besatzung in Gewahrsam. Nachdem China am 9. und 12. September die Freilassung der Gefangenen gefordert hatte, entließ Japan die Crew, der Kapitän wurde weiterhin zurück gehalten. Die Spannung kochte weiter hoch, bis China am 21. September ohne Vorwarnung alle Exporte von seltenen Erdelementen einstellte. Wegen ihres großen Exportanteils, waren die Auswirkungen weltweit spürbar. Japan beschwerte sich zwar über diese ökonomische Kriegsführung ließ den Kapitän aber dennoch frei, nach 3 weiteren Tagen. Es dauerte 1 Monat bis China den Export wieder aufnahm, und 2 Monate bis wieder nach Japan geliefert wurde.

Dieser Zwischenfall war ein Weckruf für andere Länder und so wurde wieder in die Förderung von seltenen Erdelementen investiert. In den USA ist die Mountain Pass Mine bald wieder bereit für die Produktion, in Japan arbeitet man an der Förderung von Unterwasservorkommen und auch Australien hat viele Minen wieder geöffnet. Hat China also falsch gehandelt?

Unser Modell legt diese Vermutung nahe, schließlich hatte China seine Monopolstellung schon lange Zeit, ohne dass jemand etwas dagegen unternommen hat. Die Persistence lag also sehr hoch, und die Ressource hätte sehr wahrscheinlich noch länger Bestand gehabt. Des weiteren war die Stealth sehr gering, der Engpass an seltenen Erdelementen fiel sofort aus. Zwar bestritt China jeglichen politischen Zusammenhang, die Verbindung war aber zu offensichtlich. Das wiederholte Stoppen des Exports, ist nun auch erschwert, da andere Länder nun versuchen unabhängig zu werden. Die Frage ist, ob das zurückgewinnen der Gefangenen den Einsatz einer so wertvollen Ressource wirklich wert war, wenn zukünftige Situationen wesentlich größere Gewinne bieten.

4. Ausblick in die Zukunft

4.1 Zero – Day – Exploits als Geschäft

Zu Beginn des Cyberkrieges erfuhren Firmen ihre Schwachstellen kostenlos, meistens aus dritter Hand. Sei es aus Foren oder Usermeldungen. Doch mit der steigenden Bedeutung von Exploits, begann sich ein Markt zu entwickeln, mit Preisen für einen Zero – Day – Exploit die über 100.000 US Dollar liegen.

ADOBE READER	$5,000–$30,000
MAC OSX	$20,000–$50,000
ANDROID	$30,000–$60,000
FLASH OR JAVA BROWSER PLUG-INS	$40,000–$100,000
MICROSOFT WORD	$50,000–$100,000
WINDOWS	$60,000–$120,000
FIREFOX OR SAFARI	$60,000–$150,000
CHROME OR INTERNET EXPLORER	$80,000–$200,000
IOS	$100,000–$250,000

Bild 2: Jeder Preis setzt die neueste Version an Software und einen Exklusivverkauf voraus, plus die Bedingung den Hersteller nicht mehr zu alarmieren[14]

Heutzutage stellt das einen cleveren Hacker vor schwere Entscheidungen. Er kann falls er eine Schwachstelle entdeckt, diese dem Hersteller melden, und somit sein Ansehen steigen und als Berater einsteigen. Er kann zu einer Sicherheitsfirma gehen, die ihm Geld für diese Information bietet, und damit auch die Schwachstelle beheben. Oder er wendet sich an einen Makler, der einen Deal mit der Regierung eines Landes arrangiert und somit vermutlich den größten Profit herausschlägt.(14) Deswegen verkaufen viele talentierte Hacker, die einst Microsoft oder andere Firmen warnten, die entdeckten Schwachstellen an den Höchstbietenden. Dabei gehen die Schwachstellen häufig an den Westen, denn in China sind genügend einheimische Hacker, welche die Preise drücken und in Russland ist die Kriminalität zu hoch. Fakt ist, die besten Preise erhält man in Amerika und Europa. Zwar bieten auch die Softwarehersteller selber Belohnungen an, können aber in diesen Preissegmenten nicht mithalten. So bietet Google maximal um die 3200 US – Dollar für die komplexesten Fehler in ihrer Software.[14]

In diesem Schwarzmarkt ist die USA der größte Käufer. Jedoch nutzen die USA diese Informationen nicht zur Verteidigung, sondern verfolgen die Strategie, durch diese Exploits selber ihr Potenzial im Cyberkrieg zu steigern. Laut dem ehemaligen Berater des Weißen Hauses für Cybersecurity Richard Clarke, lege die USA zu viel Wert auf ihre offensiven Möglichkeiten, was Konsumenten und Firmen in Gefahr brächte.[13] Hat also eine fremde Macht Zugriff auf die selbe Ressource, ist ein Unternehmen wie Microsoft schutzlos ausgeliefert, obwohl die Regierung theoretisch die Möglichkeit hätte Vorsorge zu treffen. Außerdem besteht die Gefahr dass nach eigenem Einsatz einer Ressource, diese schnell dupliziert und vice-versa eingesetzt werden kann, was wieder einheimische Unternehmen in Bedrängnis bringt

Dabei ist der Markt noch lange nicht gesättigt. Selbst in Software, die häufig genutzt wird und weit verbreitet ist, werden immer wieder neue Schwachstellen gefunden. Wieder dienen hier Browser als Beispiel, so wurden in Firefox 400 und in Chrome sogar 800 Lücken gefunden von 2009 – 2012. Mit der Weiterentwicklung von Software, zum Beispiel durch Patches um Schwachstellen zu beheben oder neue Features einzuführen, entstehen ständig neue Möglichkeiten Zero – Day – Exploits zu entdecken.

Mit dem steigenden Interessen von Staaten an solchen Exploits, steigen auch die Anzahl an unabhängigen Entdeckungen. Das hat laut unserem Modell mehrere Folgen. Zum einen sinkt die Persistence, da es immer wahrscheinlicher ist, dass ein neuer Hacker die Schwachstelle entdeckt und sie somit bekannt wird. Das führt dazu dass Ressourcen wesentlich eher genutzt werden, weil die lohnenswerte Grenzen sinkt. Zum anderen sinken durch steigendes Angebot selbstverständlich auch die Preise. Howard Schmidt, der wie die gleiche Position wie Richard Clarke inne hatte, nannte es naiv zu glauben dass man längere Zeit als einziger Zugriff auf einen Zero – Day – Exploit hat.

4.2 Cyberkrieg als ernste Gefahr

Die Bedeutung, die Cyberkrieg, einnimmt ist nicht zu unterschätzen. Besonders in unseren modernen Zeiten, in denen die meiste Infrastruktur aus Software besteht. Der Fall Saudi – Aramco hat gezeigt, welche globalen Auswirkungen möglich sind, sollte ein Angriff Erfolg haben. Ebenfalls besteht die Gefahr dass Terrorgruppen an besonders wertvolle Exploits gelangen, und

doi: 10.2313/NET-2015-03-1_03

diese nicht zu lange aufheben, sondern einfach darauf aus sind Schaden zu verursachen. Es existieren zwar schon Übereinkünfte zwischen den Staatengemeinschaften, diese werden aber von niemandem Ernst genommen. Notwendig ist ein allgemeines Umdenken, welche Gefahr solche Exploits bieten. Wird zum Beispiel ein Atomkraftwerk lahmgelegt, könnte dies schnell zu verheerenden Auswirkungen in einem Land führen. Die NATO hat bereits das Cooperative Cyber Defence Centre of Excellence gegründet um besser geschützt zu sein gegen solche Maßnahmen. Dieses dient zur Beratung in kritischen Fragen, versucht Forschungsarbeiten zu publizieren und arbeitet an einem rechtlichen Rahmen zur Cyberverteidigung. Eine Möglichkeit die Gefahr zu verringern wäre, sich auf eine Abrüstung für Cyberwaffen zu einigen wie es bereits mit anderem Kriegsgerät geschehen ist.

4.3 Anwendung des Modells in Spielen

Ein Modell, das von so vielen Überlegungen und Variablen abhängt ist prädestiniert für Strategiespiele. Spionage ist bereits ein oft genutztes Mittel in Spielen. Der interessante Faktor den uns das Modell bietet ist, dass man alle Ressourcen im vornherein festlegen kann und so ziemlich genau den Wert einer jeden errechnen kann. Spieler können nun extra Belohnungen erhalten je näher sie dem Idealwert kommen beim Einsatz. Das ganze ist natürlich einfacher in einem PvE Universum, da so nicht besonders auf das verhalten des Computers Rücksicht genommen werden muss. Am ehesten wäre der Einsatz in Spielen wie Civilization (Strategie - Simulationsspiel) denkbar, da dort durch das Durchschreiten verschiedener Zeitepochen viele Ressourcen denkbar sind und sich so nicht allzu schnell ein gewissen Schema einspielt.

Wesentlich komplexer wäre der Einsatz in groß angelegten Online Spielen a la EVE Online, die ebenfalls sehr simulationslastig sind. EVE Online ist ein MMORPG, welches im Weltraum stattfindet und sich auf Handel und Kampf der unterschiedlichen Fraktionen untereinander konzentriert. Das hier bereits annähernd echte politische Gebilde vorherrschen, zeigt sich in den vielen Allianz Kriegen oder auch groß angelegten Diebstählen, die es teilweise sogar in die Medien schaffen.[15] Des weiteren herrscht dort bereits ein funktionierendes Wirtschaftssystem. Ein realistisches Modell zur Einschätzung von Profit beim Hacking oder Sabotieren würde dem Spiel gewiss einen neuen Schuss Realismus verschaffen.

Eine letzte Möglichkeit wäre die Anwendung in Serious Games, um Situationen der Wirklichkeit nachzustellen. Allgemein lassen sich viele neue Ansätze finden, die dem Realismus in Spielen Auftrieb verschaffen können.

5. Zusammenfassung

Diese Arbeit sollte ein grundlegendes Verständnis über Cyberkonflikt vermittelt haben. Dieser hat schon seit längerem begonnen und sollte nicht unterschätzt werden. Das Eindringen in Schwachstellen kann sowohl zum positiven genutzt werden, wie zur Verbrechensbekämpfung oder Vorbeugung von Angriffen, gleichzeitig kann aber auch der Spieß umgedreht werden und Nationen geraten in starke Gefahr. Die Bedeutung wird in Zukunft sehr wahrscheinlich eine noch größere Rolle spielen. Die Ergebnisse aus unserem Modell zeigen, dass sowohl Persistence als auch Stealth wichtige Attribute für einen Exploit sind. Die gegensätzliche Effekte dieser zwei auf den optimalen Zeitpunkt sind auch deutlich hervorgegangen, wie in 2.2 behandelt. Genau so spielt auch die Verteilung der Einsätze eine große Rolle. Wichtig ist nicht den Fehler zu machen, seine Schwachstellen in Bezug auf das zu sehen was selber auf dem Spiel steht. Einem Angreifer reicht der derzeitige Zustand an Einsätzen vielleicht bereits aus um zuzuschlagen. Ebenso so wichtig ist es dass der Gewinn sich nicht nur aus dem direkten Nutzen ableitet, sonder auch unerwünschte Nebeneffekte auftreten können. Bei Entdeckung steigt automatisch die Wachsamkeit des Zieles und politische Konsequenzen können folgen. Herrscht Kriegszustand zwischen zwei Parteien gilt es die Risiken noch genauer abzuwägen. Dennoch ist es weiterhin schwer abzuschätzen welchen Wert die Persistence und Stealth einer Ressource besitzen. Hier kann man nur Maßstab an bereits existierenden Quellen und Beispielen nehmen. Weiterhin ist immer noch schwer zu sagen wie schnell eine Schwachstelle behoben wird. Auch hier dienen nur Studien als Maßstab.

6. Quellenangabe

[1]http://www.cyberconflict.org/blog/2013/8/1/what-are-the-costs-of-cyber-crimes-and-cyber-espionage.html

[2]http://users.ece.cmu.edu/~tdumitra/public_documents/bilge12_zero_day.pdf

[3]http://en.wikipedia.org/wiki/Conficker

[4]http://de.wikipedia.org/wiki/Saudi_Aramco

[5]http://bakerinstitute.org/files/641/

[6]http://www.darkreading.com/attacks-and-breaches/saudi-aramco-restores-network-after-shamoon-malware-attack/d/d-id/1105991?

[7]http://www.nytimes.com/2012/12/10/business/global/saudi-aramco-says-hackers-took-aim-at-its-production.html?_r=0

[8]http://www.nytimes.com/2013/05/07/world/asia/us-accuses-chinas-military-in-cyberattacks.html?pagewanted=all&_r=0

[9]Donilon T (2013) Press Briefing By National Security Advisor Tom Donilon (The White House, Washington)

[10]http://www.n-tv.de/politik/China-bestellt-US-Botschafter-ein-article12862406.html

[11]http://www.dailytech.com/World+Trade+Org+to+China+on+Rare+Earth+Metals+Stop+Breaking+the+Law/article34597.htm

[12]http://en.wikipedia.org/wiki/2010_Senkaku_boat_collision_incident

[13]http://www.reuters.com/article/2013/05/10/us-usa-cyberweapons-specialreport-idUSBRE9490EL20130510

[14]http://www.forbes.com/sites/andygreenberg/2012/03/23/shopping-for-zero-days-an-price-list-for-hackers-secret-software-exploits/

[15]http://www.t-online.de/spiele/id_19390698/milliardenraub-in-onlinespiel-eve-online-.html

[16]http://searchsecurity.techtarget.com/feature/Private-market-growing-for-zero-day-exploits-and-vulnerabilities

[17]http://m.eet.com/media/1154886/25731-electronics_industry_braces_for_rare_earth_materials_shortages_pdf.pdf

[18]http://www.pnas.org/content/111/4/1298.full.pdf+html

doi: 10.2313/NET-2015-03-1_03

Exploring DDoS Defense Mechanisms

Patrick Holl
Betreuer: Oliver Gasser
Seminar Future Internet SS2014
Lehrstuhl Netzarchitekturen und Netzdienste
Fakultaet fuer Informatik, Technische Universitaet Muenchen
Email: holl@in.tum.de

ABSTRACT

Nowadays, Distributed Denial-of-Service (DDoS) attacks are a major threat for all sizes of networks. The number of attacks against companies and institutions steadily increased over the last years. Downtime of an enterprise network usually causes financial damage. Therefore, it is important to have mechanism for DDoS defense. In this paper, various DDoS defense mechanisms are reviewed and compared with focus on rule and model based approaches. Large Botnets allow for new kinds of attacks like flash crowd simulation which mimic a huge mass of organic traffic. These kind of attacks are difficult to detect and new defense techniques are required. In order to discover new mitigation algorithms, it is necessary to understand at which layers attacks can happen. Therefore, we take a look on how attacks are classified in current research literature. In addition to the attack classification, rule and model based DDoS defense mechanisms are reviewed. For both model and rule based techniques scenarios exist where one algorithm outperforms the other one. Having this in mind, we list the advantages and drawbacks of both techniques based on insights of research literature. Emerging architectures like SDN may change the way DDoS defense is handled. Researchers are already working on algorithms that are suitable in SDN environments. The goal of this paper is to summarize current defense mechanisms and give a brief outlook on how DDoS defense could look like in the future.

Keywords

DDoS attacks, DDoS Defense, DDoS Mitigation, Algorithm comparison

1. INTRODUCTION

Denial-of-Service (DoS) attacks are a major threat for the availability of the global internet infrastructure. The main goal is to limit or even prevent intended users to access a service. In most cases, an attacker controls several compromised machines which are distributed over the internet. Such distributed attacks are also called Distributed Denial-of-Service (DDoS) attacks. Lately, attack networks with over 400,000 compromised machines were revealed [1]. A huge Botnet like this is able to cause severe availability problems even to large web services. Depending on the offered service, downtime can cause loss of revenue or other negative effects for the one who runs the service. As a consequence, defense mechanisms to detect and mitigate such DDoS attacks are necessary. DDoS defense is an active field of research but also the attackers evolve their tools and algorithms to overcome detection and mitigation. In this paper, we want to give an overview of several DDoS attacks on the one side and defense algorithms on the other side. Professional DDoS attacks often aggregate traffic from their compromised machines in a way that it looks like organic traffic from intended users. Attacks that mimic natural users can be particularly difficult to detect. Lately, new ideas emerged how to do DDoS defense in modern network architectures like Software-defined networking (SDN). However, no studies about how the proposed mechanisms work in real world environments exist nowadays.

The increasing complexity of DDoS attacks requires many-faceted defense mechanisms. Therefore, modern defense systems make use of several detection and mitigation techniques. DoS attacks can be handled in various ways, e.g. by building an infrastructure around the service which is able to survive a DDoS attack by deploying resources dynamically based on the packet load the service gets. Reactive defense mechanisms are on the other hand algorithms that try to detect and mitigate attacks at the time they occur. The reactive approaches are classified as rule or model based. Model based approaches check for traffic anomalies and rule based ones for certain patterns, e.g. in a specific packet header field.

Not all defense mechanisms are suitable for all kinds of DDoS attacks. In some cases, rule based algorithms can outperform statistical approaches, for instance, when the setup time must be very short. However, there are also scenarios where model based algorithms have advantages over rule based ones, e.g., in blocking Zero-Day DDoS attacks. Zero-Day DDoS attacks are not yet publicly known attacks.

In section 2 of this paper, DoS and DDoS attacks are defined in more detail. Section 3 gives an overview of DDoS defense mechanisms that emerged over the last years in research literature. In section 4 we compare rule based and statistical approaches and state the advantages and shortcomings of each approach. The last section 5 gives an outlook on the future of DDoS defense in SDN. SDN is an emerging technique in networking but has not yet replaced traditional architectures. DDoS defense in SDN is an active area of research right now but no studies are available yet that prove or falsify the concepts and hypotheses of the researchers.

doi: 10.2313/NET-2015-03-1_04

2. DEFINING DOS AND DDOS

To be able to mitigate DoS and DDoS attacks, respectively, it is necessary to understand the difference between both attacks. In the following section, the differences and characteristics of those kind of attacks are described. Furthermore, the basic structure of a DDoS attack is analyzed. In the last part of this section, DDoS attacks are classified based on the information of current research.

2.1 DoS vs. DDoS

The primary goal of DoS attacks is to make a service or the whole network unavailable to its intended users. To this end the DoS attack targets a network node to hinder it from processing packets that originate from legitimate requests [2].
A Distributed-Denial-of-Service (DDoS) attack can be seen as a special form of a DoS attack. In this case, distributed means the usage of multiple machines to attack the target [3]. This basic difference is visualized in Figure 1.

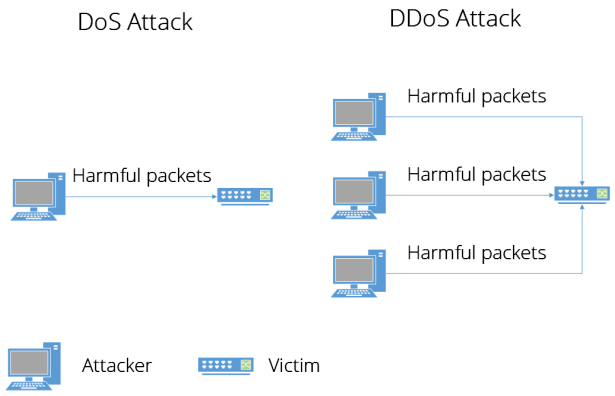


Figure 1: DoS vs. DDoS

2.2 Basic structure of a DDoS attack

Figure 1 shows only a very simplified view of a DDoS attack. The actual structure of a typical attack is more complex. Attacks that are able to take down large web services typically need several thousand compromised machines. For example, 2008 a Botnet consisting of over 400,000 machines was revealed [1]. The coordination of such a huge, distributed attack network is complex and typically done in a three layered structure as described by Kelm et al. in [4].

On the first layer is the attacker itself who controls several handlers on layer two. The handlers on layer two are used to automatically compromise and control machines to act as agents on layer three. To compromise the agents, the handlers use automated routines to find and exploit vulnerabilities. One handler can control hundreds of agents which are then used to send harmful packets to a victim. In Figure 2 we can see how the traffic, separated in control and attack parts, flows within the DDoS attack structure.

2.3 DDoS classification

To be able to develop and understand defense mechanisms, it is necessary to understand the different tiers on which a

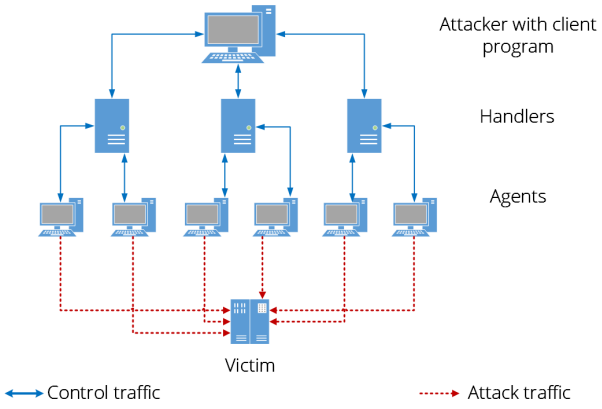


Figure 2: DDoS Control Layers

DDoS attack can happen. Several possible classifications exist nowadays which are described in detail in [5] by Douligeris et al. In the following section, we focus on the generalization of possible attack types stated in [6]:

Bandwidth-based (flood) attacks
The attacker uses its agents to send a mass amount of junk IP traffic to the victim. Consider a scenario where a website running on an arbitrary HTTP server is attacked. The webserver can only handle a certain amount of users or to be more specific, HTTP requests. If the attacker can saturate that maximum number of requests, the webserver is not able to respond to legitimate requests anymore.

Transport Layer Attacks
Protocol attacks exploit vulnerabilities and features in certain protocols. One of the most common DoS protocol attacks is TCP SYN flooding [7]. The attack exploits a weakness of the three-way handshake which is necessary to set up a TCP connection between two hosts. A valid three-way handshake consists of three messages which are sent between the client and the server. The last message is normally sent by the client and acknowledges the connection with the server. An attacker now drops the last message, meanwhile the server is waiting for the response and is – depending on its implementation – blocked for new TCP connections. Many more protocol attacks exist nowadays as described in [8] and [5].

Application Layer Attacks
Another kind of DoS attacks are targeting the application layer. An example for such a DoS attack is shown in [9] by Kulkarni where the target is the popular Apache2 webserver. Application layer attacks can target any application that is reachable via a network, e.g. expensive database requests. The underlying protocol of the applications is secondary but most commonly the HTTP protocol is used.

Actual attacks are often not easily classifiable because they exploit characteristics of more than one type. Application layer and protocol attacks often comes hand in hand with bandwidth-based attacks.

doi: 10.2313/NET-2015-03-1_04

3. DDOS DEFENSE MECHANISMS

Denial-of-Service attacks can cause severe damage on the infrastructure of the attacked victim [10]. Consider an e-commerce company which sells products online. Any downtime of the website means loss of revenue, since no legit users are able to use the service. Therefore, it is necessary to develop systems and algorithms to mitigate DDoS attacks and their impact on a service.

In the following section, we take a look at various DDoS mitigation techniques and the technical challenges that comes hand in hand.

3.1 Defense approaches

In [11], Zhang et al. categorizes DDoS defense and detection into three basic categories.

Proactive defense mechanisms

In 2002, Keromytis et al. [12] proposed a method which actually does not tackle a DDoS attack directly but built the infrastructure in a way that it will survive a DDoS attack. This implies that the attacked victim needs access to resources that can handle and survive a DDoS attack. Nowadays, such infrastructures could be called Cloud or Cloud-hosting where resources are only extended when needed. Such an infrastructure can be the only method to survive so called Zero-Day DDoS attacks, which are attacks that are not yet publicly known – and therefore no defense mechanism is available.

Reactive defense mechanisms

The concept behind reactive defense mechanisms is to mitigate or block a DDoS attack when it happens. This can be a challenging endeavor since the attack must be observable by certain patterns. An Intrusion Detection System (IDS) works as a traffic monitor and analyzer [11]. Thus means, that the DDoS defense is only as strong as the deployed IDS. Nowadays, for many DDoS attacks exist mitigation techniques for example TCP SYN Flooding [13] or ICMP Flooding [14].

Post attack analysis

The main goal of post attack analysis is to analyze an attack and find patterns in it to feed the IDS with, and on the other hand, to trace back the attacker [11]. Song et al. presented in [15] a method to trace back a spoofed IP address to its real source. However, Zhang et al. showed that it is not feasible to trace back large Botnets at the moment of the attack. One reason is that large, modern Botnets consists of thousands of agents and second one is that the global internet is too big that all administrators can collaborate to exchange trace back information.

All of the named defense approaches can be combined and applied together. For example, selective blackholing (see section 3.2) as a reactive approach to mitigate the attack itself and a dynamic *cloud* that can supply additional computation power on demand. The reactive mechanism helps to mitigate the attack in such a way that less additional resources are necessary to handle and survive the attack.

3.2 Selective Blackholing

A classical blackholing approach can be used to block traffic which is destined to a certain victim [16]. Therefore, all packets from a certain IP address that causes high traffic are routed to a so called null route. Depending on the geographical region of the source address, the traffic could theoretically also originate from legitimate request, e.g. due to some advertisement. Having this in mind, one major drawback of this approach is that not only malicious but also legitimate traffic is filtered out. In order to tackle these shortcomings, selective blackholing emerged.

Selective DDoS blackholing is a two-step process with the goal of sending all DoS related packets to a static route defined on the edge network routers to drop them [17]. In the first step of the process, all edge routers are initialized with a so called *blackhole* destination. All packets forwarded to this destination are separated from regular traffic and usually dropped. In the second step, the BGP routers in the network use the specified blackhole destination to forward packets and instruct the service provider when certain conditions are met [18]. A possible condition would be for example, a malformed packet or an IP address which is known for being an agent in a large Botnet.

In 2014, Snijders presented a selective blackholing approach which also takes the geographical scope into account [18]. Consider the following scenario: A web shop that only sells and ships products to German addresses. Most likely the customers of this shop will access it with a *German* IP given by their ISP. A large Botnet is usually distributed over several countries because the agents are (in most cases) infected by automated routines that exploit vulnerabilities in the system. A selective blackhole that takes the geographical scope of German IP addresses into account can now be used to block traffic outside this scope. Figure 3 illustrates this case.

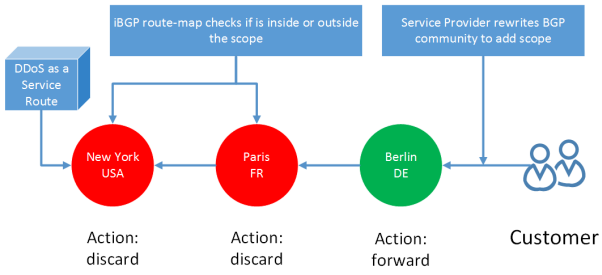


Figure 3: Selective blackholing, discard outside Germany

doi: 10.2313/NET-2015-03-1_04

Router name	Continent ID	Country code	Metro ID	Latitude, Longitude
r1.tky.jp	3	392	46	35.65671„139.80342
sjo.us	1	840	29	37.44569,-122.16111
dal.us	1	840	30	32.80096„-96.81962

Table 1: Router geographical locations table [18]

Snijders used the following four rules to illustrate his algorithm [18]:

* discard traffic sourced outside 'this' country (5580:664)
* discard traffic sourced outside 'this' continent (5580:660)
* discard traffic sourced outside a 1000 km radius from 'here' (5580:663)
* discard traffic sourced outside a 2500 km radius from 'here' (5580:662)

According to Snijders, the *this* and *here* keywords are points that refer to the point where a customer interconnects with the service [18]. The two-numbered code in brackets stands for a autonomous system and an action. For instance, *5880:664* means that the AS with the code 5580 wants to discard all traffic outside the country where a customer interconnects with the service. This action is represented by the second code 664. A router can only set packets on a null route if they have a route map where they can check whether the destination is outside the geographical scope or not. Therefore, a table which contains the geographic location of the routers is necessary. Table 1 shows how such a database could look like. In this table, column one states the name of the routers, column two to four are geographic indicators. For instance, the *Metro ID* with the value 46 represents a number code for Tokyo, Japan. The last column contains geographic coordinates for distance calculation.

Packets with source addresses that are routed through routers outside the geographical location defined by the service, can now be set on a null route and discarded.

However, selective blackholing as described above is not able to block a DDoS attack completely. When the attack traffic originates from an IP address which is within the defined scope, the traffic would still reach its target. But since only traffic from its main target group reaches the service, the attack can be heavily mitigated and the service can continue its business. Anyhow, scope based selective blackholing also has some shortcomings which we will discuss in section 4 in more detail.

3.3 Statistical Approaches

Statistical approaches are based on the assumption that DDoS attack traffic shows anomalies in the entropy and frequency of selected packet attributes. In 2003, Feinstein et al. proposed an algorithm to detect DDoS attacks by measuring statistical properties in packet headers at different points in the network [19].

A mandatory basis of every statistical detector is the model on which it is built on. For instance, the model can be generated based on a certain number of legitimate requests within a defined time range. Assume a web service provider that logs all consecutive packets from 9.00PM to 9.15PM for one month. After that month the service provider is able to build a model (for the given time range) that contains information about the distribution of the source IP addresses. In the second month, incoming packets can be checked against the model and classified as forward or drop.

3.3.1 Entropy of consecutive packets

One method proposed by Feinstein et al. is based on the entropy comparison of consecutive packet samples to identify changes in their randomness [19]. Information entropy is the average amount of information in each sample and defined as follows, where H is the entropy, n the number of symbols and p_i the occurrence probability of symbol i:

$$H = -\sum_{i=1}^{n} p_i \log_2 p_i \tag{1}$$

The source IP address is one field in the packet header that can be used to identify deviations in the randomness in comparison with legitimate requests. Depending on the number of agents, the attacker only has a limited number of IP addresses he can use. Therefore, attacks can be identified if the number of unique IP addresses has a wide variance from the legit samples.

One shortcoming of this technique is that an attacker who knows how the algorithm works is able to break it by slowly forging packets until they match the right entropy levels. However, this is not a trivial task since multiple detectors can be chained together which makes it harder to break through all of them.

Detection Quality
Feinstein et al. evaluated their proposed algorithm by simulating a DDoS attack based on an excerpt of 1,000,000 packets from the NZIX dataset[19]. They decomposed the packets into 75% legitimate and 25% DDoS traffic. TCP SYN flooding is used as an attack. Therefore, the 25% attack traffic consists of TCP SYN flooding packets. The packets used for the attack were numbered from 700,000 to 800,000. Figure 4 shows the result that the researchers got from the attack by applying an entropy model using the source IP addresses as stated above. The calculated entropy values are mapped to the y-axis and the packet count, in thousand, is mapped to the x-axis.

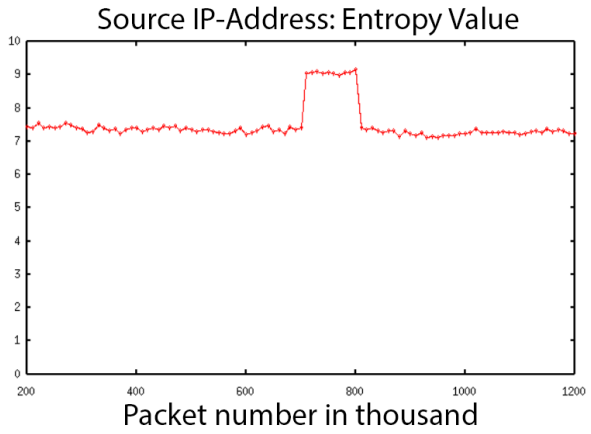


Figure 4: DDoS entropy result (Source: [19])

3.3.2 *Chi-Square Statistic*
If the number of measurement values is small, e.g., a binary value which is either 1 or 0, the entropy might not be sufficient enough to calculate reliable thresholds. As a consequence, Feinstein et al. also made use of the Chi-Square statistic. Therefore, Feinstein looked at the TCP SYN flag distribution of consecutive incoming packets. In some scenarios, the source address distribution is not an appropriate base. For instance, when Network Address Translation is used to map several source addresses onto one unique address. The TCP SYN flag is a discrete value, it is either 1 for set or 0 for unset. Pearson's chi-square Test is a suitable method to compare the distribution of discrete measurement values [19].

Pearson's chi-squared test is defined as follows, where B is the number of cells (e.g. 2 for the TCP SYN flag values), N_i the number of packets where the corresponding values occur and n_i is the expected number of packets under a normal distribution.

$$\chi^2 = \sum_{i=1}^{B} \frac{(N_i - n_i)^2}{n_i} \tag{2}$$

Detection Quality
Feinstein et al. used the same setup for the chi-square test as for the entropy test [19]. Regarding the DDoS detection quality, both techniques offer the same accuracy (see Figures 4 and 5). The only difference is that the thresholds are different to classify a packet as a DDoS packet. In this case, all packets with a χ^2 value over around 1,500 can be considered as harmful.

3.3.3 *Conclusion*
Model based statistical approaches have one major advantage over rule based techniques like selective blackholing. They allow detection for Zero-Day DDoS attacks. However, it can be a challenging task to set up an appropriate model in practice. For instance, a DDoS detector for a website with a constantly growing user base would classify legitimate requests as attack if the underlying model is static. Therefore, the model has to be constantly updated which can be tricky since the update must happen when the server is not under attack. Otherwise the model is falsified and not suitable for attack detection anymore. Another issue of model based techniques is to find the right thresholds that classifies anomalies. If the thresholds are set too low, many intended users are blocked but if the thresholds are too high, the DDoS attack can cause more damage.

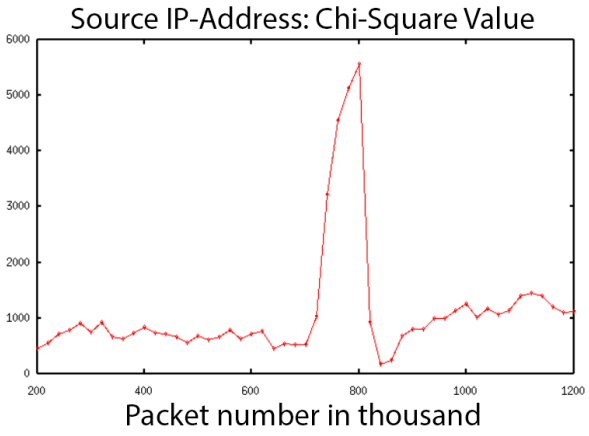


Figure 5: DDoS entropy result (Source: [19])

3.4 Challenges
As we have seen in section 3.3.3, it is challenging to provide a model based DDoS detection for an evolving website. In fact, many more technical challenges exist to detect and mitigate actual DDoS attacks. In the following section, we will discuss what DDoS attacks makes so hard to mitigate nowadays.

3.4.1 *Size of the Botnet*
Attackers who control large Botnets with tens of thousands of compromised machines can go beyond flooding attacks like TCP SYN flooding. A modern DDoS technique is to mimic a flash crowd. In this case, a flash crowd is a huge number of intended users that is accessing a web service due to some suddenly appeared popularity. For example, a blog article which is linked on the front page of a major

news site can cause high traffic for the blog. The detection of such an attack is not trivial since both legitimate and malicious requests do not differ in their content but only in their intention.

3.4.2 Abnormal traffic detection

Despite the fact that we can use statistical approaches as stated in Section 3.3 to detect anomalies in a consecutive packet flow. It is still a very challenging task to identify packets forged by an attacker which make use of several obfuscation techniques. Flash crowd imitation is one example for a technique which is hard to defend. Detection on the one hand is relatively easy since the number of requests just increases up to an abnormal state. But defense on the other hand is hard because the attacked service provider might block a real flash crowd and permanently disgruntle intended users with it. These kind of attack is very depending on the resources an attacker can use for his attack. A Botnet with a high geographical distribution is, for instance, difficult to mitigate with selective blackholing since the attacker can use bots that are within the geographical fence.

3.4.3 Long-term attacks

On the long-term, an attacker may harm a service more if he uses attacks that not completely prevents intended users from using it but increases its respond times. The goal of these attacks is to utilize a service to its capacity. They are both, hard to detect and hard to mitigate, because at the first appearance a service looks like in its default state. The only difference is that due to its high utilization a service has a longer respond time. Services that are time critical can take severe damage from such attacks. For instance, Amazon.com[1], which is one of the largest e-commerce provider on the planet, estimated a loss of 1,600,000 USD in sales if its page load time would increase by one second for the period of one year [20].

3.4.4 Large-scale testing

It is a game of cat-and-mouse between the ones who develop attacks and the ones who develop defense algorithms. The development of DDoS defense mechanisms is a complicated task because it is hard to test the developed algorithms in real-world scenarios. One reason for that is the lack of large-scale testbeds, another one is that it is not safe to perform experiments within the actual internet infrastructure [8]. Nowadays, common methods are small-scale test setups and simulations as we have seen in section 3.3.

On a commercial level, several service providers exist which offer large-scale tests. Not all of them are reliable, e.g. if they don't do any site owner verification. Such providers fall in the category of DDoS-as-Service as described in [21]. Reliable companies that offer large-scale tests don't do this for free usually. Instead they charge a price that relates with the size of the attack. As a consequence, large-scale testing can be very expensive.

[1] http://amazon.com

4. COMPARISON OF DDOS DEFENSE TECHNIQUES

In the last section we have seen several DDoS defense techniques like rule based or model based filtering. Depending on the concrete attack, one algorithm can outperform the other one. The different defense techniques can also be combined in order to increase the DDoS mitigation level. In this section we want to discuss and compare the advantages and disadvantages of the proposed techniques in section 3.

4.0.5 Advantages of rule based filtering

As described in [18], rule based filtering can be an effective way to mitigate certain DDoS attacks. In comparison with statistical approaches, rule based filters don't require a model to detect attacks. Building accurate distribution models for parameters like the source addresses can be a difficult and time consuming task. Rule based filters on the other side require much less setup time. They can start working immediately after they are setup i.e. as soon as the rules are made. Furthermore, rule based filtering allows a detection rate of 100% if a certain attack happens for which rules are already defined. In addition to that, the number of false-positive results is (depending on the rules) very low. The maintenance of a rule based filter usually requires less effort than a model based one. One reason for this is that rule based filtering is independent of the number of packets and the traffic. In comparison with that, statistical models must be constantly updated in order to fit the parameter distribution of the legitimate requests.

4.0.6 Disadvantages of rule based filtering

Application layer attacks as described in section 2 usually exploit vulnerabilities or software design mechanisms. In order to block such attacks, the filter rules must match the attack pattern. However, for unknown vulnerabilities no such rules can be defined which means that Zero-Day DDoS attacks cannot be blocked by rule based filtering. The only possibility to mitigate Zero-Day DDoS attacks is to set up generic rules like selective blackholing [18]. In this case, selective blackholing can be seen as a generic rule because packets are not further analyzed but blocked based on their geographic origin only. Nowadays, many different types of DDoS attacks exist which target the victim on different layers, see section 2. Usually, different DDoS attacks require different detection rules which results in a large repository of rules that is required to block those different kind of attacks. In comparison with that, model based approaches identify harmful packets on their abnormal parameter distribution - without having different rules for any single attack.

4.0.7 Advantages of model based filtering

One major advantage of model based approaches is that Zero-Day DDoS attacks can be mitigated. This is because traffic streams which have a high deviation with respect to the model built from the legitimate requests are flagged as potentially harmful. In addition to that, one model can be used to mitigate different kinds of attacks. For instance, an arbitrary attack where the attacker sends consecutive packets which have an abnormal source address distribution. In this case, it is not necessary to further analyze the payload of the packets themselves.

4.0.8 *Disadvantages of model based filtering*

Statistical approaches suffer from the so called cold-start problem. It takes time to build an appropriate model before attacks can be detected and mitigated. In addition to that, the models have to be constantly updated in order prevent a high rate of false-positives when the website is evolving (i.e. getting more traffic by intended users) over time. As a consequence, it can take time until the models reflect the actual situation. However, most websites are not evolving in a speed which invalidates a model very fast. Another difficulty in defining model based filters is finding the right thresholds. On the one hand, if the threshold is set too low, legitimate requests are blocked and therefore intended users, on the other hand, if the threshold is set too high, malicious requests are not blocked and therefore harm the service.

4.0.9 *Conclusion*

Both rule based and statistical approaches have advantages over the other technique. Depending on the concrete scenario, one technique can outperform the other one. Services that require a more robust detection for Zero-Day DDoS attacks should use a model based mitigation technique. Services that require a very short setup time should primarily apply a rule based filter. Since there exists a plethora of different kinds of DDoS attacks, one detection algorithm alone may be not sufficient enough. Therefore, it is possible to chain and combine several defense techniques like selective blackholing and a source address distribution model with concrete thresholds. Consider the following scenario:

A German online shop which has 8600 customers in total and 8500 of them live in Germany. A majority of 95% of all orders is shipped to a German address.

In this scenario, Germany is the main market and responsible for most parts of the revenue. A selective backholing algorithm that blocks all packets from source addresses outside Germany could be the first line of defense. Attacks from a globally distributed Botnet are severely mitigated in this case. For attacks launched by German hosts, the source address distribution of consecutive packets can be used as a second line of defense and to block abnormal packet streams. All in all we can say that a plethora of different kinds of attacks and threats like Zero-Day DDoS attacks require the combination of various defense mechanisms to take advantage of their specific strengths.

5. FUTURE TRENDS

In this section we will discuss future trends in networking and how they possibly affect DDoS defense and attack mechanisms. Vykopal et al. represents the hypothesis that SDN is ideal for distributed DDoS detection and mitigation [22].

The traditional network architecture which is based on TCP/IP is now over 20 years old but still the major technique for transmitting packets in a network. Due to trends like Cloud and the Internet of Things, the demands on network technology is constantly increasing. In this case, Cloud stands for centralized, outsourced service providers that offer disk space and applications as a service. One technique which is emerging over the last years is SDN or Software-defined networking. Figure 6 shows a schema of a SDN architecture with its three layers. SDN allows Cloud providers to easily separate the traffic from their customers into flows.

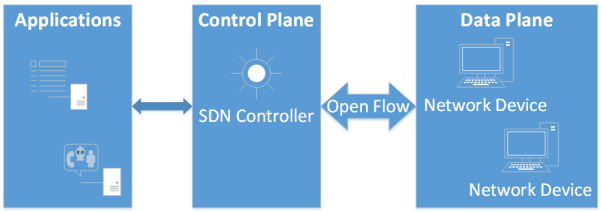


Figure 6: Software Defined Networking architecture schema

One major point that brings Vykopal to his hypothesis is that a SDN is flow based. The data plane and the control plane are separated from each other. A switch in a traditional network contains both control and data planes whereas switches in a SDN only contain the data plane. The packets are forwarded based on the entries of the flow table managed by the SDN controller. Since the traffic is organized in flows through the SDN, there is a possibility that attack flows can be identified by certain patterns. Another point of Vykopal which makes SDN ideal for him concerning DDoS mitigation is that there is a central point of knowledge, i.e. the SDN controller. Once a malicious flow is identified within the network, the controller can block or blackhole the respective flow.

At the time of this paper no studies that prove or falsify the hypothesis of Vykopal et al. were available. However the researchers will focus on three main research questions and try to answer them over the next three years [22]. The first one is a generic investigation of the differences that SDN brings to traditional networks and its monitoring. In a second step, the researchers try to explore the specific vulnerabilities in the data and control plan of a SDN. Furthermore, Vykopal et al. wants to use that discovered knowledge afterwards to find out how DDoS attacks in Software Defined Networks can be optimally mitigated.

6. CONCLUSION

DDoS attacks are one of the largest threats for the global internet nowadays. The attacks can be used to slow or even shut down large network infrastructures. Therefore, DDoS defense is a necessary task to ensure the availability of the internet. In this paper, we tried to give an overview of various kinds of DDoS attacks and how they can be detected and mitigated. Steadily, new kinds of attacks on the one hand and new defense mechanisms on the other hand are discovered. Having this in mind, it is mandatory to constantly update attack patterns and signatures for detection and mitigation purpose. Another very important point is to develop algorithms which are able to mitigate Zero-Day DDoS attacks, so that at least the main intended user group is still able to access the service. In 2014, Snijders presented a technique called selective blackholing which we discussed in section 3 [18]. Depending on the concrete scenario, selective blackholing can be a very strong defense against DDoS

attacks. However, it still has some drawbacks, e.g. if the users are not locally concentrated. As we have seen in section 4, selective blackholing could be combined with a model based algorithm in order to harden its defense abilities. In the future, multiple lines of defense can play a much more important role. This is mainly because DDoS attacks are evolving by getting more complex and resources behind it. However, SDN can dramatically change the way DDoS defense is done. For now, we don't have any major studies on this and it is an ongoing field of research as we have seen in section 5. As a consequence, the superiority of SDN in DDoS defense remains speculation at the time of this paper. In conclusion, further research in SDN DDoS defense is necessary.

7. REFERENCES

[1] "Spam on rise after brief reprieve." `http://news.bbc.co.uk/2/hi/technology/7749835.stm`. Accessed: 2014-08-30.

[2] E. Y. Chen, "Detecting dos attacks on sip systems," in *VoIP Management and Security, 2006. 1st IEEE Workshop on*, pp. 53–58, IEEE, 2006.

[3] A. Asosheh and N. Ramezani, "A comprehensive taxonomy of ddos attacks and defense mechanism applying in a smart classification," *WSEAS Transactions on Computers*, vol. 7, no. 7, pp. 281–290, 2008.

[4] K. Möller and S. Kelm, "Distributed denial-of-service angriffe (ddos)," *Datenschutz und Datensicherheit*, vol. 24, no. 5, pp. 292–293, 2000.

[5] C. Douligeris and A. Mitrokotsa, "Ddos attacks and defense mechanisms: classification and state-of-the-art," *Computer Networks*, vol. 44, no. 5, pp. 643–666, 2004.

[6] NSFOCUS, "Introduction to ddos attack," 2004.

[7] W. M. Eddy, "Tcp syn flooding attacks and common mitigations," 2007.

[8] J. Mirkovic and P. Reiher, "A taxonomy of ddos attack and ddos defense mechanisms," *ACM SIGCOMM Computer Communication Review*, vol. 34, no. 2, pp. 39–53, 2004.

[9] P. Kulkarni, *Responsive System for DDoS Attack against Apache Web Server*. PhD thesis, NATIONAL INSTITUTE OF TECHNOLOGY KARNATAKA, 2010.

[10] M. Sachdeva, G. Singh, K. Kumar, and K. Singh, "Measuring impact of ddos attacks on web services," 2010.

[11] G. Zhang and M. Parashar, "Cooperative defence against ddos attacks," *Journal of Research and Practice in Information Technology*, vol. 38, no. 1, pp. 69–84, 2006.

[12] A. D. Keromytis, V. Misra, and D. Rubenstein, "Using overlays to improve network security," in *ITCom 2002: The Convergence of Information Technologies and Communications*, pp. 245–254, International Society for Optics and Photonics, 2002.

[13] H. Wang, D. Zhang, and K. G. Shin, "Detecting syn flooding attacks," in *INFOCOM 2002. Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, vol. 3, pp. 1530–1539, IEEE, 2002.

[14] L. Limwiwatkul and A. Rungsawang, "Distributed denial of service detection using tcp/ip header and traffic measurement analysis," in *Communications and Information Technology, 2004. ISCIT 2004. IEEE International Symposium on*, vol. 1, pp. 605–610, IEEE, 2004.

[15] D. X. Song and A. Perrig, "Advanced and authenticated marking schemes for ip traceback," in *INFOCOM 2001. Twentieth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, vol. 2, pp. 878–886, IEEE, 2001.

[16] M. Caesar and J. Rexford, "Bgp routing policies in isp networks," *Network, IEEE*, vol. 19, no. 6, pp. 5–11, 2005.

[17] J. Van der Merwe, A. Cepleanu, K. D'Souza, B. Freeman, A. Greenberg, D. Knight, R. McMillan, D. Moloney, J. Mulligan, H. Nguyen, *et al.*, "Dynamic connectivity management with an intelligent route service control point," in *Proceedings of the 2006 SIGCOMM workshop on Internet network management*, pp. 29–34, ACM, 2006.

[18] J. Snijders, "Ddos damage control cheap and effective," 2005.

[19] L. Feinstein, D. Schnackenberg, R. Balupari, and D. Kindred, "Statistical approaches to ddos attack detection and response," in *DARPA Information Survivability Conference and Exposition, 2003. Proceedings*, vol. 1, pp. 303–314, IEEE, 2003.

[20] "Slow websites cost retailers billions." `http://mashable.com/2012/11/22/slow-websites/`. Accessed: 2014-10-24.

[21] J. J. Santanna and A. Sperotto, "Characterizing and mitigating the ddos-as-a-service phenomenon," in *Monitoring and Securing Virtualized Networks and Services*, pp. 74–78, Springer, 2014.

[22] M. Vizváry and J. Vykopal, "Future of ddos attacks mitigation in software defined networks," in *Monitoring and Securing Virtualized Networks and Services*, pp. 123–127, Springer, 2014.

What's New in the Linux Network Stack?

Lukas M. Märdian
Advisors: Paul Emmerich, Daniel Raumer
Seminar Future Internet, Winter Term 14/15
Chair for Network Architectures and Services
Department for Computer Science, Technische Universität München
Email: maerdian@in.tum.de

ABSTRACT
In this paper, interesting features of the Linux kernel's network stack are analyzed, which were introduced during the development cycles from Linux v3.7 to Linux v3.16. Special attention is given to the low-latency device polling, introduced in Linux v3.11, the netfilter's SYNPROXY target, introduced in Linux v3.12 and the new Nftables framework, introduced in Linux v3.13. At the end a trend is presented, which shows the direction in which the Linux network stack is evolving.

Keywords
Linux, network, packet processing, SYN proxy, JIT compiler, firewall, low-latency, Berkeley Packet Filter

1. INTRODUCTION
The Linux kernel is a fast moving and always changing piece of free software. Having a very mature network stack, Linux is deployed widely, especially to drive and manage the ever growing and changing world wide web. It is sometimes hard to follow the newest developments and discussions regarding the Linux kernel, so this paper gives an analysis of the current state of the Linux kernel's network stack and presents some interesting features, which were introduced in the development cycles from Linux v3.7 to v3.16.

Chapter 2 describes the low-latency device polling feature, which was introduced in Linux v3.11. Chapter 3 gives an overview of the addition of a SYN proxy to the Netfilter subsystem, introduced in Linux v3.12. The new packed filtering framework *Nftables*, successively replacing *Iptables*, is introduced in Chapter 4. In Chapter 5 the trend of the Linux kernel's network evolution is discussed and finally a conclusion is presented in Chapter 6.

2. LOW-LATENCY DEVICE POLLING
In the Linux kernel v3.11 Eliezer Tamir et al. introduced a low-latency polling mechanism for network devices. The classical way how the Linux kernel handles its network devices, by using the New API (NAPI), provides a good trade off between efficiency and latency of the packet processing. Still, some users with more specific demands, such as the finance sector with their high-frequency trading systems or scientific research with high-performance computing, need to reach the lowest latency possible, even on high traffic network devices. This demands are not possible to reach with NAPI. [3]

2.1 Interrupts vs. Polling
Usually, the Linux kernel handles network devices by using the so called New API (NAPI), which uses interrupt mitigation techniques, in order to reduce the overhead of context switches: On low traffic network devices everything works as expected, the CPU is interrupted whenever a new packet arrives at the network interface. This gives a low latency in the processing of arriving packets, but also introduces some overhead, because the CPU has to switch its context to process the interrupt handler. Therefore, if a certain amount of packets per second arrives at a specific network device, the NAPI switches to polling mode for that high traffic device. In polling mode the interrupts are disabled and the network stack polls the device in regular intervals. It can be expected that new packets arrive between two polls on a high traffic network interface. Thus, polling for new data is more efficient than having the CPU interrupted and switching its context on every arriving packet. Polling a network device does not provide the lowest packet processing latency, though, but is throughput optimized and runs with a foreseeable and uniform work load. [1]

2.2 Low-latency Polling
In order to make network packets reach the network stack and user space as fast as possible, the low-latency device polling mechanism was introduced, so users of the network stack (applications) can poll for new packets, whenever they are ready to process new data. Even though polling is involved, this technique provides a lower latency than reenabling the per-packet interrupts, because on a high traffic network device, there would be hundreds or thousands of packet-interrupts per second. Handling all of them would introduce a larger latency than polling within very small intervals. [2, 3]

Technically, this is realized by a new function call, named `ndo_busy_poll()`, defined in `include/linux/netdevice.h`, which can be implemented by network device drivers [4]. In this function call the device drivers are supposed to poll the network hardware for new packets and return them immediately. If no packets are available at the moment, the drivers are supposed to *busy wait* for new arriving packets until a timeout is reached, as defined (in μs) by `sysctl.net.core.busy_read` and `sysctl.net.core.busy_poll`. The default timeout value is set to 0, which means busy waiting is disabled and the driver should poll only once. For latency critical applications a polling timeout of 50μs is recommended. [2, 3, 7]

doi: 10.2313/NET-2015-03-1_05

On devices, whose drivers implement this low-latency polling function call, e.g. Intel's *ixgbe* driver [5], the network stack will poll the hardware for new network data at certain situations, e.g. when it is instructed to do so by a user space application via the `poll()` system call. This way the user space can ask for new network packets, whenever it is ready to process new data. The driver will then either directly flush the new network packets, which arrived since the last poll, or (if no packet arrived) it will poll the network device in a busy waiting loop and flush new data, arriving before the timeout. Another situation where the network stack will issue the low-latency polling function call is in the `read()` system call: When a user space application tries to read new data from a socket but there are no packets in the queue, the network device will be polled in a busy waiting loop until the timeout is reached. On high traffic devices, the driver will most likely be able to return new data to the user space as a response to its `read()` call. [2, 3]

With this low-latency polling mechanism, an latency improvement of about 30% is possible (2.5μs within an UDP connection and 2.2μs within a TCP connection), introducing just a minimally increased CPU load on a typical system, using two Intel Xeon E5-2690 CPUs, X520 optical NICs and the *Netperf* benchmark (c.f. Figure 1). [7]

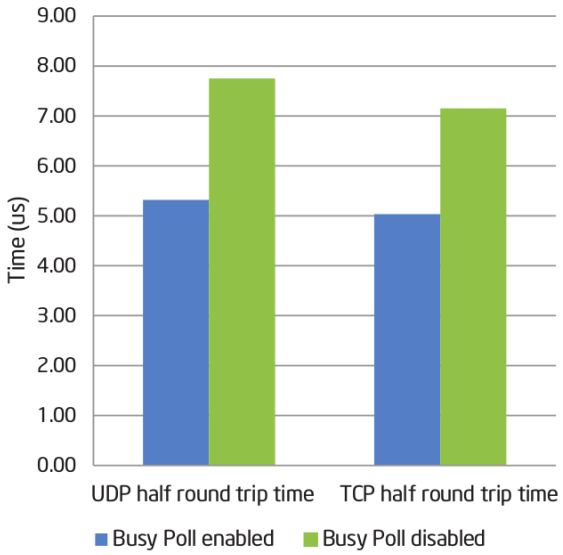


Figure 1: Netperf Latency results [7]

3. NETFILTER: SYNPROXY TARGET

Introduced in the Linux kernel v3.12 by Patrick McHardy et al. was a *SYNPROXY* target for the kernel's internal packet filtering framework, called *Netfilter*. This implements the concept of a stateful firewall, where connection attempts are filtered very early in the network stack, before they reach their destination. So they can be handled before they are tracked by a data structure, called *transmission control block* (TCB), in the filtering subsystem (*conntrack*).

3.1 SYN-Flooding Attacks

In order to establish a TCP connection between a server and a client, a three way handshake is used: The client sends a SYN packet to the server to request a connection, then the server responds with a SYN/ACK packet to confirm the client's request. If everything is fine, the client answers with an ACK packet and the connection is established. This TCP three way handshake can be attacked by a SYN flooding attack, which is a commonly used Distributed Denial-of-Service (DDoS) attack. In this attack, the attacker sends many SYN requests with spoofed IPs to the server, using a botnet. For each of those faked requests the server needs to initialize a TCB data structure, respond with a SYN/ACK packet and wait for an ACK from the client. In case of a DDoS attack the ACKs will never arrive, as the senders IP addresses are not real. Still, the server needs to store the TCBs for a certain amount of time, which will fill up its TCB buffer and leads to a situation where real (i.e. non-attack) requests cannot be handled anymore and thus the server is rendered unreachable. [8, 9] Reducing the timeout of the transmission control blocks, which is 5 seconds in the default case (initial retransmission timeout (1 sec) * `tcp_synack_retries` (5), c.f. `man tcp(7)`, [10]), or increasing the TCB buffer will help to handle such situations. Depending on the attacker's botnet size, the server will be out of service anyway, though. In order to defend such DDoS attacks, countermeasures, such as SYN cookies and SYN proxies have been invented, which are presented in the following.

3.2 SYN-Cookies

One countermeasure to defend a SYN flooding attack is the use of so called SYN cookies. The modern type of SYN cookies was introduced in 1996 by D. J. Bernstein et al. [11]. The idea is to compute the cryptographic hash of the characteristic data of the connection, like the sender's and receiver's IP addresses and the ports used. This data is joined with a secret key, which is only known to the server, using a hashing algorithm, such as MD5 or SHA. The resulting hash is called SYN cookie and is set to be the sequence number (SQN) of the SYN/ACK TCP packet. If the SYN/ACK reaches a real client, which responds with a real ACK message, the cookie can be reconstructed from that message's SQN field. Thus the server does not need to store a TCB for each SYN request, but instead just needs to check the cookie of arriving ACK packets. If the SYN cookie contained in the SQN field of the arrived message is the same as the hash, which the server can reconstruct from the secret key and the connection's characteristics, the server can be sure that he send a SYN/ACK message to that client before. Using this method, SYN flooding can be circumvented, as there is no TCB queue, which could be flooded. There are some drawbacks in using SYN cookies, though. For example the server needs to have enough computing power to compute SYN cookies for all arriving SYN packets in real time and the firewall needs to know the server's TCP options, to determine the connection characteristics. [9]

3.3 SYN-Proxies

A SYN proxy is an entity in the same network as the server, which ought to be protected. Its purpose is to filter and/or load balance incoming SYN requests, using different methods and technologies (e.g. SYN cookies). A common method

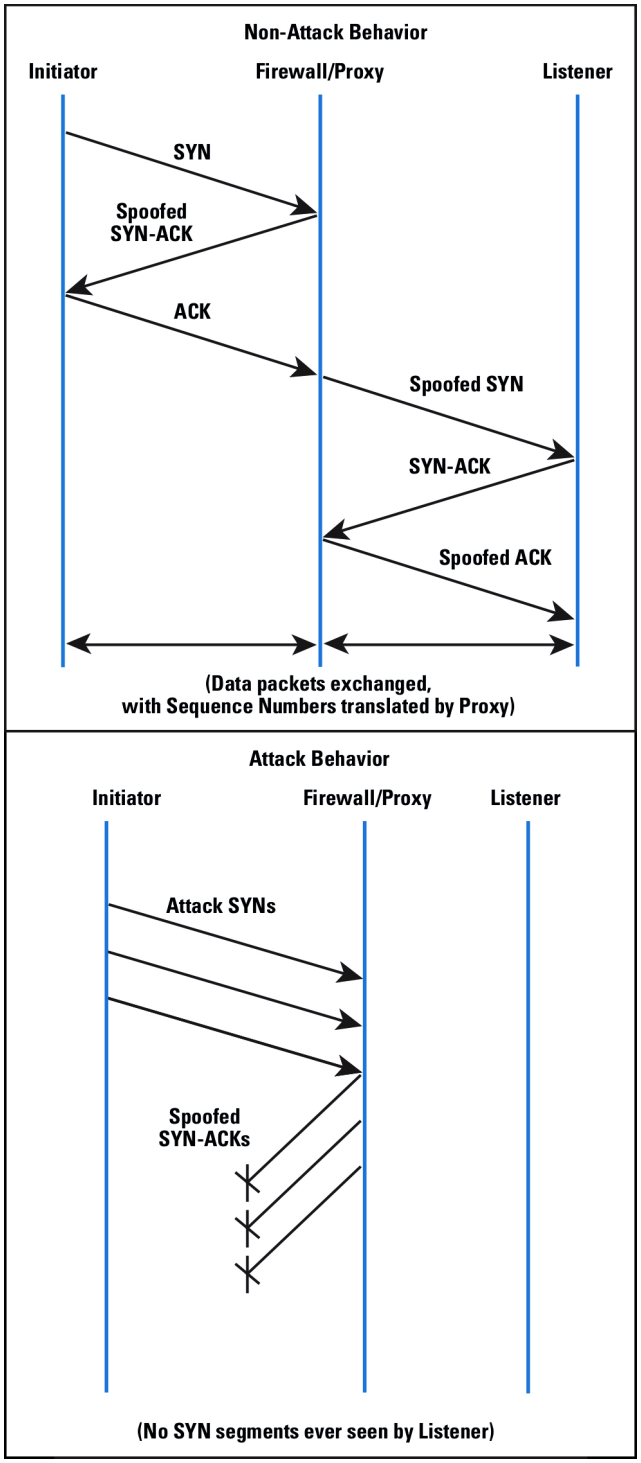


Figure 2: SYN-ACK spoofing Proxy [14]

of such a stateful firewall is the spoofing of SYN/ACK packets to the client (cf. Figure 2), in order to avoid invalid connection attempts to reach the real server and open a connection, including a TCB. [14, 12]

In Linux v3.12 and beyond the new SYNPROXY target can be setup as shown in the following:

1. Don't track incoming SYN requests:
```
iptables -t raw -I PREROUTING -i $DEV -p tcp \
-m tcp --syn --dport $PORT -j CT --notrack
```
2. Mark unknown ACK packets as invalid:
```
/sbin/sysctl -w \
net/netfilter/nf_conntrack_tcp_loose=0
```
3. Handle untracked (SYN) and invalid (ACK) packets, using the SYN proxy:
```
iptables -A INPUT -i $DEV -p tcp -m tcp \
--dport $PORT -m state --state \
INVALID,UNTRACKED -j SYNPROXY --sack-perm \
--timestamp --wscale 7 --mss 1460
```
4. Drop the remaining invalid (SYN/ACK) packets:
```
iptables -A INPUT -i $DEV -p tcp -m tcp \
--dport $PORT -m state --state INVALID -j DROP
```
5. Enable TCP timestamps, needed for SYN cookies:
```
/sbin/sysctl -w net/ipv4/tcp_timestamps=1
```

Using this setup and a little bit of *conntrack* tuning, it was shown, that the performance of a server during a SYN flooding attack can be increased by the order of one magnitude, using this *Netfilter* SYNPROXY target [12]. This is a nice improvement, but of course using this stateful firewall comes at a cost: Due to the extra layer, which the SYN proxy introduces, the connection establishment phase will take longer, leading to an increased latency. Also TCP connection parameters and characteristics, used by the SYN proxy, must match the servers TCP settings. These settings have to be set in the SYNPROXY module manually, leading to a configuration overhead and a new source of errors. To fix the configuration problem, the authors have already proposed a solution where the connection characteristics can be detected automatically, by forwarding the first connection request for normal processing to the server and sniffing the settings from its response [13].

4. NFTABLES

Nftables is a dynamic firewall system, which is based upon a bytecode interpreter. The project was initially started in 2009 by Patrick McHardy but discontinued, due to a lack of interest. In October 2012 the Netfilter maintainer Pablo Neira Ayuso announced to revitalize Nftables, based upon McHardy's work. With his additional ideas he could attract more developers and together they were able to integrate Nftables as a new feature in Linux v3.13. [15]

4.1 Iptables, Ip6tables, Arptables, Ebtables

Prior to the general purpose firewall system *Nftables* several other, protocol specific solutions existed in the Linux kernel, namely *Iptables* to filter IPv4 connections, *Ip6tables* to filter IPv6 connections, *Arptables* to filter ARP connections and *Ebtables* to filter ethernet bridging connections. Those static solutions had some drawbacks, such as tightly coupled data structures, which got passed back and forth between the kernel and user space. This made the kernel's implementation of those filters quite inflexible and optimizations to the structures very hard to implement. Furthermore, the ruleset for the filters was represented as one big chunk of binary data, which made it impossible to incrementally add

new rules to the firewall system. The solution to this problem was, that the firewall management tools dumped the firewall's current state to a file, modified the whole chunk of data (added, removed, altered rules) and injected the file as a whole back into the firewall system. This approach works but is problematic, because it has a quadratic complexity for incremental changes. Also, by injecting the new rule set, the firewall will loose its current state, which makes it unable to continue tracking the currently open connections, so it has to start over. [17]

4.2 Virtual Machine

The concept behind the dynamic *Nftables* system, which helps to overcome the above mentioned problems of static and protocol specific firewall systems, are based upon a virtual machine. All the static filtering modules, used by *Iptables* and the other protocol filters, are replaced by a single kernel module, providing "a simple virtual machine [...] that is able to execute bytecode to inspect a network packet and make decisions on how that packet should be handled" [15]. The idea of this approach is inspired by the *Berkeley Packet Filter* (BPF) virtual machine. The virtual machine on its own is not able to do any packet filtering, instead it is dependent on small bytecode programs, which describe how a specific package should be handled. Those small, individual programs get compiled by Nftables' corresponding management tools in the user space, e.g. the command-line utility `nft`, and can incrementally be put into the interpreter at runtime. [16]

One of the biggest benefits of this virtual machine approach is a massive reduction in complexity. By replacing four filtering systems for different protocols with a single, universal system, a lot of code can be removed. This leads to less problems, because duplicated code is removed and the remaining common is reviewed by more people. By moving the filtering logic out of the Linux kernel into the firewall management tools, while just keeping a small virtual machine to execute the bytecode, provided by the management tools, the complexity of the filtering logic is reduced as well.

In order to stay compatible with the old firewall management tools, such as `iptables` and `ip6tables`, Nftables provides a compatibility layer, which is integrated in the Iptables project and enables long time users of Iptables to easily migrate their old firewall rules (using old syntax). Internally the new bytecode for the Netfilter virtual machine will be compiled. However, new users are encouraged to use the new `nft` command-line utility to manage their firewall in Linux kernels of version v3.13 or beyond. [16, 15]

4.3 New possibilities of Bytecode filtering

In contrast to the old firewall system, the new dynamic system, using its small bytecode programs, is much more flexible and provides quite some improvements: Each rule change, which is represented as a small bytecode program, can be performed atomically during the runtime of the system and without interfering with the general state of the firewall. Also, open connections can be kept open, if not requested differently be the new rules. This atomic replacement of rules does speed up changes in firewalls with big rule sets quite a bit, as it is not needed anymore to dump, modify and replay the whole state of the system, but instead just the single, wanted modification can be executed. In addition to the speedup it also helps to avoid race conditions, which could occur during the rule set change in the old system. [15]

Another benefit is, that new matching types (i.e. characteristics of a packet to filter for) can be added easily to a bytecode program. Whereas the old system used to depend on an extra kernel module for each matching rule, which could be set by the management tools. This led to a large amount of over 100 modules. Getting new matches/modules into the kernel took much longer than writing a simple program, which can be injected into the firewall at runtime. Furthermore, handling all the matching in small programs, compiled by user space tools, makes it possible to have the rule set optimized. Using generic compiler optimizations, faster execution in the in-kernel virtual machine can be achieved, e.g. by automatically removing duplicated or unreachable rules, which the firewall administrator did not think of. [17]

Using the `libnftnl` library, provided by the Nftables project, it is easily possible to write new and improved firewall management tools, too. This library enables those user space tools also to listen to changes in the firewall system and notify the applications in real time about the current state. [17] One tool leveraging this library is `nft`, the main firewall management tool, provided by the Nftables project. It has a similar functionality as the old `iptables` tool but tries to be more intuitive, by using natural language to describe the different filters instead of lots of configuration switches. For example if somebody wants to drop all IPv4 HTTP traffic, this can be established as follows: [18]

1. Setup an iptables like chain, using the ipv4-filter file, provided by Nftables:
   `nft -f files/nftables/ipv4-filter`

2. Drop all incoming TCP packets with the destination of port 80 (HTTP):
   `nft add rule ip filter input tcp dport 80 drop`

5. LINUX NETWORKING TREND

In this final chapter, a general trend is shown, which can be observed in the development of the Linux kernel's network stack: It can be seen that the kernel developers introduce more and more dynamic features such as virtual machines, bytecode interpreters and JIT-compilers into the kernel, which help to abstract certain features and move the complexity into the userspace.

5.1 Berkeley Packet Filter VM

The Berkeley Packet Filter (BPF) has for very long been part of the Linux kernel. It is a tool, which can filter network packets on a low level, in order to function as a per-application firewall, forward only relevant packets to the user space and allow the tracing of network traffic, using tools such as `tcpdump` [19]. Simple examples of the filter's internal workings can be found in the original paper by McCanne et al. 1993, e.g. a small BPF program, which loads a packet's Ethernet protocol type field at offset 12 and accepts the packet if it is of type *IP* or rejects it otherwise:

doi: 10.2313/NET-2015-03-1_05

Listing 1: BPF program: IP filter [23]

```
    ldh     [12]
    jeq     #ETHERTYPE_IP, L1, L2
L1: ret     #TRUE
L2: ret     #0
```

Since the release of Linux kernel v3.0, the BPF has been improved continuously. It started with the introduction of a just-in-time (JIT) compiler for the filter, by Eric Dumazet, which enabled the Linux kernel to translate the virtual machine instructions to assembly code on the x86_64 architecture in real time and continued with other performance improvements and functional extensions in Linux kernel v3.15 and v3.16. [19, 20]

In recent Linux kernel versions (v3.15, v3.16), the BPF virtual machine has been extended from having a very simple and network specific architecture with just a few registers and capabilities, to being more of a general purpose filtering system, whose capabilities can be mapped pretty close to modern hardware. The instructions of this virtual machine, as can be seen in Listing 1, are usually mapped 1:1 to real assembly instructions of the underlying hardware architecture. [19, 21]

5.2 Dynamic Firewall Systems

The concept of dynamic filtering systems on different levels of abstraction is another development trend in the Linux kernel. As discussed before, there is Nftables, which has its filtering rules created and optimized dynamically at user space level, using tool such as `nft`. These rules, which implement the filters, can then be fed into the firewall system dynamically at runtime and are being processed by a virtual machine in the kernel, in a dynamic manner (c.f. Chapter 4).

In addition to the general purpose firewall *Nftables*, which aims to protect a Linux system from the outside world and regulate the incoming and outgoing network traffic, there is the *Berkeley Packet Filter* (BPF), too. It can function as a per-application firewall and can be used for system internal packet processing, e.g. to reduce the network traffic, which is directed to a specific application. The BPF is implemented as a dynamic bytecode interpreter, too. (cf. chapter 5.1)

This evolution from static tools, such as `iptables` or the early, static variant of the *Berkeley Packet Filter*, to more dynamic and abstract tools, is characteristic for the direction of development in the current versions of the Linux kernel. A remaining question is, why those quite similar tools are still separated and not merged into one universal solution for dynamic packet filtering. This is because the tools come from different backgrounds and as of today none of them is yet fully able to replace all the others. McHardy explained, why they did not build *Nftables* upon an existing solution, such as the BPF virtual machine: "A very important feature, one that is missing from all other filters that are built similar in the kernel (like BPF, TC u32 filter, ...), is reconstruction of high level constructs from the representation within the kernel. TC u32 for example allows you to specify 'ip daddr X', but when dumping the filter rules it will just display an offset and length." [22]

As the abstraction level continues to rise, and the firewall systems are not just used for packet filtering any more, but also for general Linux kernel tracing, it is plausible that the different virtual machines in the Linux kernel are combined into a single, general purpose interpreter in the future.

5.3 Future possibilities

Watching the current trend of developments in the Linux kernel, a little outlook how the kernel's network stack might evolve in the future is presented now: The inclusion of dynamic bytecode interpreters into the Linux kernel is a popular way to increase the level of abstraction in kernel development and also to reduce the complexity. Right now, there are several separate virtual machines in the kernel, which is not an optimal solution, because some common code needs to be implemented in the same (or minimally different) ways in all of the implementations. That is also why the kernel developers rejected the inclusion of yet another virtual machine named *Ktap* (used for kernel tracing) in favor of the BPF's tracing capabilities. [24]

In general, the BPF is a very well known virtual machine, as it is in the kernel for a very long time already and got considerable performance and functional improvements in the recent Linux kernel releases. With its split into the *classic BPF* variant, providing a legacy interface to the classic BPF network filtering functionality, and the new *internal BPF* variant, which is a generalized virtual machine, including a JIT compiler. This JIT compiler has seen various performance improvements, by optimizing its commands to use architecture specific assembly code. In the future such optimizations are likely to be continued, especially for newer architectures, such as ARM64. [21]

With the Berkeley Packet Filter's architecture, it is already possible to use it for network packet filtering and also for more general tasks, such as syscall tracing, as used by the Linux kernel's *secure computing* (seccomp) system, similar to what *Ktap* wanted to achieve [21]. Some functionality is still missing in the BPF, though. For example *Ktap* wanted to enable the possibility to use filters, supplied by user space applications and so does *Nftables* with its rule sets, generated in user space, too. Another thing, which needs to be implemented in the BPF, is the possibility to reconstruct high level data structures, i.e. reverse the optimizations, which have been done by the JIT compiler, as this is one of the reasons why *Nftables* was not build upon the BPF (c.f. Chapter 5.2).

If those and potentially some other drawbacks would be improved, the Berkeley Packet Filter could become the single, general purpose and highly optimized virtual machine in the Linux kernel and all the other tracking, tracing and filtering systems could build upon it. This would be a large benefit to Linux developers and users, as the sharing of code leads to less potential problems and a single spot to apply optimizations in, which will in turn benefit all systems.

6. CONCLUSION

In conclusion, the developments in the Linux kernel's network stack keep pace with the fast and always changing Internet. This global network has very versatile demands,

which the Linux kernel tries to adopt to, be it from a security, performance or complexity perspective.

With the addition of a low-latency device polling mechanism in Linux v3.11, the kernel can now easier be used in scenarios where very low network latencies are critical, such as high performance computing or high frequency trading. Before the inclusion of this feature, the companies who needed low latencies usually implemented their own low-latency network stack in the user space, in order to bypass the kernel. With the low-latency device polling in place, those companies can now work together with the Linux community, to continuously improve the low-latency network stack in a single place, saving a lot of development resources. Benchmarks by Intel [7] show an improvement in latency of about 30% for the TCP and UDP protocols and a use case where lots of small network packets with low-latency demands are send.

In terms of security, the Linux kernel was improved by the introduction of stateful and dynamic firewall additions. The SYN proxy, introduced in Linux v3.12, adds an extra stage to the firewall, where new network connection attempts are tracked, using SYN cookies. Only if the connection was successful, they get forwarded to the rest of the network stack, which reduces the overhead in case of a SYN flooding DDoS attack and enables a server to handle ten times more packets, with just a small increase in packet latency [12]. The *Nftables* dynamic firewall system on the other hand, introduced in the Linux kernel v3.13, improves the kernel's network stack with the new possibilities of bytecode filtering, where the filters are not statically coded into kernel modules, but rather the rules are compiled and optimized to small bytecode programs in user space. Those small programs are then executed in an in-kernel virtual machine at runtime. This way the management of the firewall system is much more flexible and can dynamically adopt to the changing demands of the Internet.

All in all the developments evolve into a direction, where more systems can be controlled from outside the kernel, by user space applications. Also, the controlling can be done in a dynamic manner, be it by the activation of a low-latency path for network packets, the redirection of network traffic to a SYN proxy or the execution of bytecode programs. In the future, the different packet processing systems in the Linux kernel might be merged into a single, abstract system, which is able to handle lots of tasks by executing code, provided by the user space in a dynamic manner. Such a system would not only be able to handle network related tasks, but could also deal with other tracking, tracing and filtering tasks.

7. REFERENCES

[1] J. Hadi Salim, R. Olsson, A. Kuznetsov: *Beyond Softnet*, In Proceedings of the 5th Annual Linux Showcase & Conference, pages 165-172, 2001

[2] J. Corbet: *Low-latency Ethernet device polling*, In Linux Weekly News, Eklektix Inc., May 2013, `https://lwn.net/Articles/551284/`

[3] J. Brandeburg: *A way towards Lower Latency and Jitter*, At Linux Plumbers Conference, August 2012

[4] E. Tamir: *net: add low latency socket poll*, June 2013, `http://git.kernel.org/cgit/linux/kernel/git/torvalds/linux.git/commit/?id=060212928670593`

[5] E. Tamir: *ixgbe: add support for ndo_ll_poll*, June 2013, `http://git.kernel.org/cgit/linux/kernel/git/torvalds/linux.git/commit/?id=5a85e737f30c`

[6] E. Tamir: *net: low latency Ethernet device polling*, May 2013, `https://lkml.org/lkml/2013/5/19/20`

[7] J. Cummings, E. Tamir: *Open Source Kernel Enhancements for Low Latency Sockets using Busy Poll*, October 2013

[8] W. M. Eddy: *RFC-4987: TCP SYN Flooding Attacks and Common Mitigations*, August 2007, `http://www.ietf.org/rfc/rfc4987.txt`

[9] J. Bongertz: *Tatort Internet: Nach uns die SYN-Flut*, In c't 15/2011, 2011, `http://heise.de/-1285780`

[10] Paxson, et al.: *RFC-6298: Computing TCP's Retransmission Timer*, June 2011, `http://tools.ietf.org/html/rfc6298`

[11] D. J. Bernstein, E. Schenk: *SYN cookies*, `http://cr.yp.to/syncookies.html`

[12] J. D. Brouer: *DDoS protection, Using Netfilter/iptables*, At DevConf.cz, February 2014

[13] J. D. Brouer: *RFE: Synproxy: auto detect TCP options*, January 2014, `https://bugzilla.redhat.com/1059679`

[14] W. M. Eddy: *Defenses Against TCP SYN Flooding Attacks*, In The Internet Protocol Journal - Volume 9, Number 4, Cisco Press, December 2006

[15] J. Corbet: *The return of nftables*, In Linux Weekly News, Eklektix Inc., August 2013, `https://lwn.net/Articles/564095/`

[16] E. Leblond: *Nftables, what motivations and what solutions*, At Kernel Recipes, September 2013

[17] P. McHardy: *nftables - a successor to iptables, ip6tables, ebtables and arptables*, At Netfilter Workshop (NFWS), 2008

[18] E. Leblond: *Nftables quick howto*, February 2014, `https://home.regit.org/netfilter-en/nftables-quick-howto/`

[19] J. Corbet: *A JIT for packet filters*, In Linux Weekly News, Eklektix Inc., April 2011, `https://lwn.net/Articles/437981/`

[20] T. Leemhuis: *Kernel-Log – Was 3.0 bringt (1): Netzwerk*, In Heise Open Source, Heise Zeitschriften Verlag, 2011, `http://heise.de/-1257064`

[21] J. Corbet: *BPF: the universal in-kernel virtual machine*, In Linux Weekly News, Eklektix Inc., May 2014, `https://lwn.net/Articles/599755/`

[22] P. McHardy: *nftables*, August 2008, `http://web.archive.org/web/20081003040938/people.netfilter.org/kaber/weblog/2008/08/20/`

[23] S. McCanne, V. Jacobson: *The BSD Packet Filter: A New Architecture for User-level Packet Capture*, In Proceedings of the USENIX Winter 1993 Conference (USENIX), pages 259-270, January 1993

[24] J. Corbet: *Ktap or BPF?*, In Linux Weekly News, Eklektix Inc., April 2014, `https://lwn.net/Articles/595565/`

Internet science-Creating better browser warnings

Sepideh Mesbah
Advisor: Dr. Heiko Niedermayer
Seminar Future Internet WS1415
Chair for Network Architectures and Services
Department of Informatics, Technical University of Munich
Email:sepideh.mesbah@tum.de

ABSTRACT

The number of internet users is increasing everyday, the number of security threats is growing as well. Browser warnings try to avoid users from being defrauded. They warn the users about the possibility of a threat, but it is always up to the user to decide whether to heed or ignore the warning. One main issue is the overwhelming amount of security warnings that each user might face, which makes it hard for the user to distinguish between serious or trivial threats. Better warnings can be created by considering items such as: how to design the display of warnings that affects the user's attention; how to involve social psychological factors in designing the warnings to have an impact on the user's decision; or realizing when to present a warning message and etc. The aim of this seminar paper is to first present the reasons why the users ignore a browser warning or turn it off and afterwards to discuss some recommendations for creating more effective warnings.

Keywords

Warnings, Malware, Social, Psychology, Behaviorism, Security, Phishing

1. INTRODUCTION

While surfing the internet, the user might visit an infectious website. Browsers like Google Chrome and Firefox will try to stop the user by showing a warning message, but it is always up to the user to proceed or return to a previous page. This indicates that the role of the user should not be ignored by security practitioners.

There are three kinds of browser warnings: **malware** warning which appears when a website intends to damage a computer or steal information, **phishing** happens when the user is sent to a fake website instead of the real one and **SSL** warning pop ups when an invalid security certificate is identified.

Every day internet users are targeted by viruses, malwares, worms, phishing, etc. Users who do not pay attention to the warnings might believe that they are able to distinguish a real threat from a fake one. They extremely trust their ability or might think they have nothing to lose and are less susceptible. Stealing is not always about money, it can be the user′s information. The victims do not consider the ability of the scammers who can steal their information [14]. The scammers can use the victims personal information and create a bank account, buy or rent a property, etc.

In May and June 2013, a study[1] analyzed more than 25 million warning screens in Google chrome and Firefox to find the percentage of users which heed web browser security warnings. Users faced a by-passable browser warning given the option to click trough the warning by 'choosing proceed anyway' in Google chrome or 'understood the risk' in Firefox. The authors implemented some metrics in browsers to count the number of times that users saw a warning but clicked through without paying attention to it.

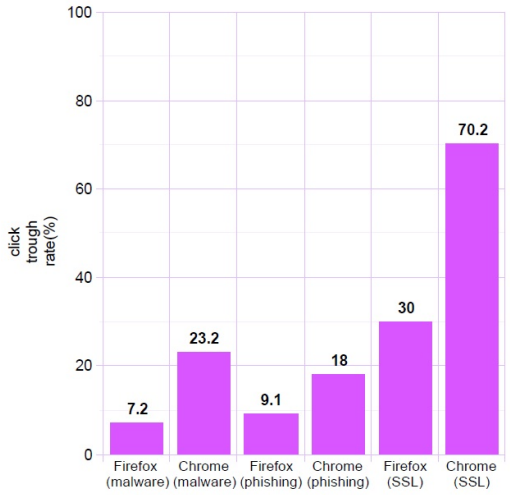


Figure 1: Click through rate (number of ignored to number of shown warnings) for Firefox and Google chrome malware, phishing and SSL warnings

As the results are shown in Figure 1, user behavior changes across different warning mechanism designs, hence more effective security warnings can be created in practice.

In this paper our main research is about how to create less but effective warnings that can attract the users attention. The paper is organized as follows: first, we present some reasons why the users ignore the warnings. Next, we focus on how to create more effective warning and will present several approaches which will be followed by a conclusion.

2. REASONS FOR TURNING OFF BROWSER WARNINGS

doi: 10.2313/NET-2015-03-1_06

In order to create better warnings, the reasons for ignoring or turning off the security warnings have to be discovered. Some individuals will ignore a warning in any way without any reason. They generally click through the security warning without looking through it. Some others think that warnings are just related to windows users and other operating system users can feel safe.

In this section we want to discuss several possible reasons for turning off browser warnings. Figure 2 is an example of a malware warning in Google Chrome.

2.1 Trust in automation

Automation has its own problems. Some users do not trust automated systems, they can not rely on it and would rather make their own decision. On the other hand some individuals trust automated systems incorrectly. For example they extremely trust their anti virus applications and think that the application can protect them from all kind of malwares, thus will download whatever they want.

The association between the users and automation can be defined with words misuse and disuse. Misuse means that people trust an automated system inappropriately which may lead them to fail. Disuse means that individuals do not trust an automated system and will ignore its ability[5]. In both situations the users do not have enough knowledge about the automated systems and may lead them to make wrong decisions.

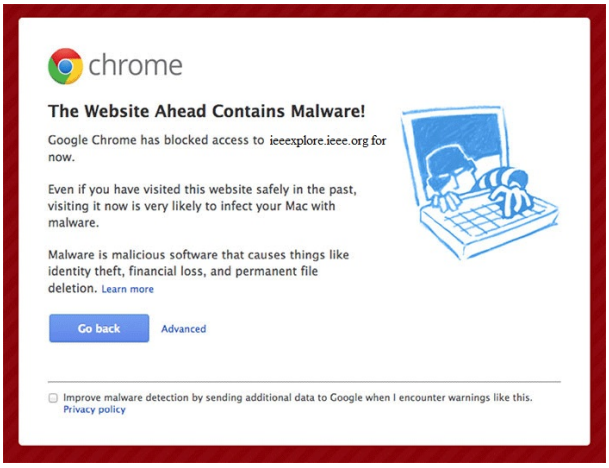


Figure 2: Example of a malware warning in Google Chrome

2.2 Not understand

Not understanding the meaning of a security warning is another reason for ignoring it. Individuals who do not understand the concept of the warning will ignore it easily. For example when users do not have enough information about the words SSL or Phishing, they will ignore it without considering the negative consequences[3].

2.3 Habituation

When the user observes a warning multiple times she might get used to it. After several times perceiving the same warning, the user gets confused with the similar looks and can not distinguish the serious alarms. User's attention decreases, thus ignores the exception message without even reading it once.

2.4 False positives

It is always difficult for the users to distinguish between a real and serious security warning from a trivial one. Analysis in [6] showed that various users didn't heed the warnings since they previously faced several false alarms. This means that they saw a warning message which tried to stop the user's operation, but when they ignored it, it later appeared to pose no threat. In this case, the users think they can identify the security risks on their own. For example browsers might give a false SSL alarm about expired security certificate of a website. This kind of alarms can be meaningless like if the computer's clock is set incorrectly, which makes the security certificate look expired.

2.5 Hassle

Some individuals are lazy and think heeding to the warning is waste of time, so will ignore it quickly. Another view is economic perspective. Herley [8] described that the likelihood of a serious attack happening is relatively low, compared to the cost of effort of reading a warning message, checking the URL for detecting phishing threats, spending time to choose strong passwords, etc.

2.6 Trusting high-reputation websites

Users might heed warnings about the websites they visit for the first time, but they won't pay attention to the warnings that appear for a safe high reputation website which the user visited before[7].

A recent study [7] analyzed around four million different Chrome malware warning effects. As it is shown in Figure 3, the users are twice as likely to ignore a visited website which was stored in their browser's history. Individuals trust high-reputation websites, thus will not heed to the warnings.

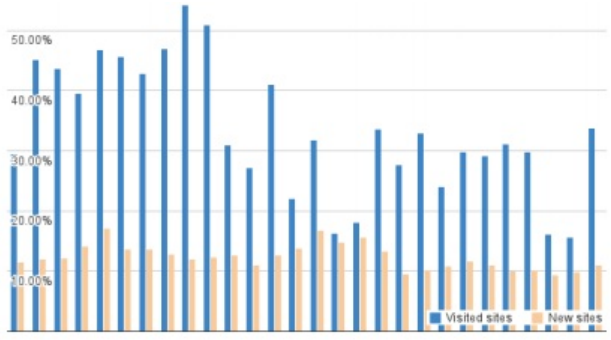


Figure 3: Y indicates the click through rate, for visited web sites (blue), or new sites (red). Each point in x axis is a day. 28 days in January 2014 [7].

3. CREATING EFFECTIVE WARNINGS

When a user visits a suspicious website, the browser will present a warning message to the user. Although the last decision is always made by the user, effective warnings can

doi: 10.2313/NET-2015-03-1_06

be designed in a way that can prevent the user from being in a hazardous situation. In this section first we will discuss when a warning should be used. Next the focus will be on the presentation of warnings, afterwards the social psychological factors will be discussed which showed to have a great impact on creating effective warnings.

3.1 When should a browser warning be used

As mentioned in [15], one of the main issues in security warnings is the habituation. The user will try to ignore the messages she has seen several times without reading it. Being aware about when a browser warning should be shown is important. Figure 4 shows a graph for risk assessment. Risk can have two features: The impact it may have and the probability that it might occur.
Three different zones are presented in Figure 4. The first zone indicates that the impact of the risk is low, so in this case it is better to not bother the user, and it is not necessary to send a security warning. In the second zone, the impact of risk is high thus the browser should block the user's action without sending any warnings. Only in the third zone it is required to ask the user to make a decision. When the impact of the risk is neither low nor high, based on the probability of risk occurrence, a warning message should be shown to the user and ask her to choose between ignoring or heeding to the alarm[15].

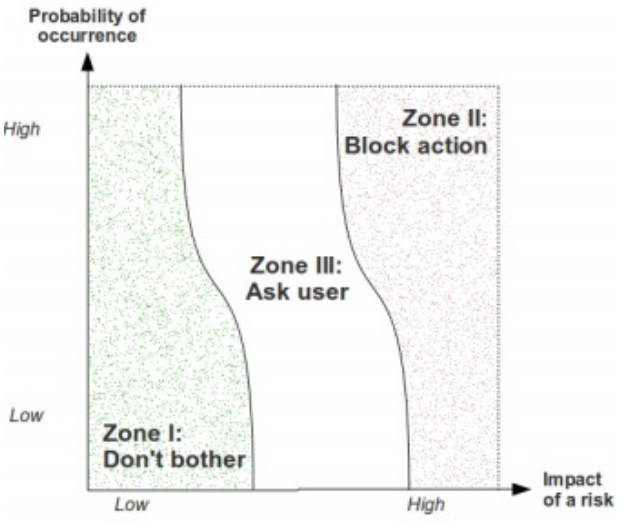


Figure 4: Risk assesment [15]

3.2 Active warnings

A good step towards creating effective warning is the usage of active warnings. New browsers use active indicators instead of the passive one and force interaction with the user. The warning gives the user choices and recommends the best option, but it is always up to the user to either heed or ignore the suggestion.

A laboratory study [2] was conducted to examine the effectiveness of active warnings. The authors used the Communication Human Information Processing Model (C-HIP) of Wogalter[1] to determine the reasons an indicator is ineffective. The C-HIP model delivers a warning message to the receiver, the receiver verifies five processing steps and the goal is to identify if the warning can change the user behavior. 60 users participated in the study. They were asked to make a purchase from Ebay or Amazon. After finishing the payment they had to check email for purchase confirmation which was a phishing message sent by the examiner. During the whole process the users were asked to think loudly, and after the experiment they had to fill out a survey.

3.2.1 Results

The reaction of the users was recorded as shown in the Figure 5.

Condition Name	Sample Size	Saw Warning	Read Warning	Recognized Warning	Understood Meaning	Understood Choices
Firefox	20	20	13	4	17	19
Active IE	20	19	10	10	10	12
Passive IE	10	8	3	5	3	5

Figure 5: The number of participants for different conditions [2]

79% of participants heeded the active warnings, but for passive warnings only one user payed attention to the warning. The results of the processing steps of the (C-HIP) are as follows:

Attention Switch and Maintenance: Active warnings interrupt the user's task and forces her to notice the indicators, but due to keystrokes users may never notice passive warnings. Warnings should be effective in a way that it can get the user's attention. Around 55% of the participant claimed to read at least one of the phishing messages completely. 19 participants stated they recognized the message. The users assumed the message is not serious since they had seen it before for trusted websites, thus ignored the message.

Comprehension/Memory: This part is to find out if user understands the meaning of the indicators. As shown in Figure 5, most Firefox users claimed they understood the meaning

Attitudes/Beliefs: The authors [2] asked the users about their attitudes and belief and how it affected their perception. The answers proved that there is a strong correlation between trust and obeying the warnings. Most of the users stated that since it gave them the option of still proceeding to the website, they thought it couldn't be that serious. Another significant correlation was between knowing the meaning of phishing and paying attention to the message. Having information about phishing made the users obey warnings.

Motivation: Overall 31 participants heeded the warning message. The motivation of the participants behavior was that they thought about the risks they might face and they wanted to feel safe. But the rest were unaware about the risks.

3.2.2 *Suggestions*

The authors of [2] presented the following suggestions:

- Interrupt users primary task and force her to notice the warning and take an action by active warnings
- Recommend a clear option which is the best choice for the user.
- If an indicator is not read by the users, then the warning should take the recommended action. This means in active warnings if the users close the message without reading it it should prevent the user from visiting the website.
- Indicators must prevent habituation. The more serious warnings should be designed different from less serious ones. In this instance, the users won't recognize and thus will pay attention to the message.
- The warning should be designed in a way that can draw inappropriate trust away from the user, so that the user won't trust for example their anti virus and heed the warning.

3.3 Warning design guidelines

Based on the work of [15] some general guidelines for the design of warnings will be discussed in this section.

Describe the risk clearly: An indicator should be designed to protect the user from being in an unsafe situation. Every warning should clarify the risk the user might face, the consequences of not heeding to it, and options for avoiding the risk.

Be concise and accurate: A warning should be brief but accurate. It should avoid long, technical and offensive text. Technical terms should be replaced with words that are easy to understand for users. At the same time the warning should include enough information that the user can perceive the risk in simple words without being oblivious to it.

Offer meaningful options: Indicators should contain two or more choices and it should suggest the best option for the user. Moreover, instead of using options like 'Ok' or 'Cancel' for disregarding the warnings, it is better to use a clear option like 'ignore this warning'. Over and above that, the location of the recommended option must be above all the other choices.

Follow a consistent layout: Figure 6 shows a suggested layout for warning messages.

- As shown in Figure 6 critical warnings should not have a close button at the upper right corner of the message, to force the user to read the content.
- The indicator should contain an icon for showing the seriousness of the warning message.
- The primary text of the warning should be a clear single sentence that can convey the importance of the message

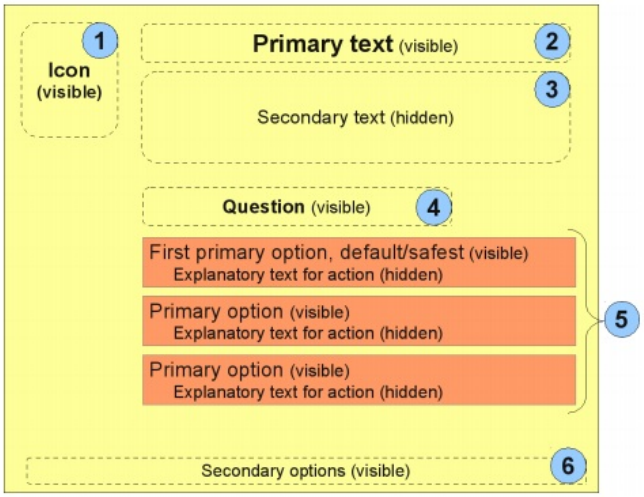


Figure 6: Warning design guideline [15]

- The warning should include secondary text for giving more information to the user but it is better to be hidden and be presented if the users clicks the more information option.
- A question should be asked about the action the user wants to take.
- Several options should be provided, and the recommended option should be the first one.
- It is good to use secondary option at the bottom of the message like 'help' which does not respond directly to the question asked from the user

In previous sections the presentation of warnings was discussed. The next section will focus on some social psychological factors used by scammers.

3.4 Social psychological factors

The goal is to create less but more effective warnings in order to improve risk communication. The social psychological factors discussed in [3] are: influence of authority, social influence, concrete threats and vague threats, which will be defined as follows:

3.4.1 *Influence of authority:*

Scammers are able to defraud users using the influence of authority. The scammers will act in the role of a trusted authority figure. Victims will trust requests from these scammers, since individuals tend to agree to request from authority figures. For example Murphy [9] showed that when the users trust the tax authorities, their willingness to pay taxes will increase. Another example [11] would be that when individuals receive emails from an ostensible doctor which suggests some drugs, they will trust the suggestion since they believe in doctors generally. In this instance, better warning can be created using influence of authority, which leads the users to heed to the warnings presented by a trusted authority figure.

doi: 10.2313/NET-2015-03-1_06

3.4.2 Social influence:

Social influence is another factor in social psychology, and it happens when individual thoughts and actions are affected by other people in the society. Being susceptible to social influence is one of the main features of peoples. Design and fashion in a community is a clear picture of social influence[12]. In marketing, the costumer will buy the item that the seller has suggested to her, even if is not her preferred item[13]. A person tends to commit more crimes if she finds out that the other members of the community also comply with committing crimes [10]. In this case, social influence can be considered for creating better warnings. The individuals which are more susceptible to social influence will comply to the request from other people from the society, for example Facebook friends, and will heed or ignore warnings.

3.4.3 Concrete and vague threats:

[16] showed that individuals which had already an information about the fraud, or individuals who tried to probe the request sent by the scammer, were less likely to be scammed. Individuals are likely to feel safe and be away from risky situations. Warning messages should be created in a way that present clear information about the negative consequences. Using concrete threats compared to vague ones, helps the individuals to gain more information about the frauds, thus will lead them to pay attention to the warnings.

3.4.4 Study:

500 users participated in the survey recruited via Amazon Turk. Five different conditions (warning) were presented to the user shown in Figure 7.

Condition	Text
Control	Control text has been taken from Google Chrome anti-malware warning as of June 2013.[a]
Authority	The site you were about to visit has been reported and confirmed by our security team to contain malware. We strongly encourage you to avoid visiting this page.
Social Influence	The site you were about to visit contains software that can damage your computer. The scammers operating this site have been known to operate on individuals from your local area. Some of your friends might have already been scammed. Please, do not continue to this site.
Concrete Threat	The site you are about to visit has been confirmed to contain software that poses a significant risk to you, with no tangible benefit. It would try to infect your computer with malware designed to steal your bank account and credit card details in order to defraud you.
Vague Threats	We have blocked your access to this page. It is possible that it might contain software that might harm your computer. Please close this tab and continue elsewhere.

Figure 7: Five different warning texts [3]

3.4.5 Result:

The research [3] showed that concrete threats had the most significant effect on users behavior. In general when users get a clear understanding about the threat and risky situation they can decide better and will heed to the warning. Appeal to authority was another factor which had influenced the users decision. When individuals receive the message from a trusted authority figure they will accept it easier, thus will pay attention to it. Another factor that had an impact on users was social influence. Individuals would click through a warning if their friends told them it is safe.

4. CONCLUSION

In this paper first several reasons were presented why the users turned off the browser warnings. Mistrust or trust in automation led the users to make wrong decisions in different situations. Misunderstanding the concept of the warning made the users to ignore or turn off the warnings. Receiving bunch of false positive warnings which later appeared to pose no threat, was another reason for users to don't heed to warning. Some other users think heeding to the warning is waste of time. Also, users won't pay attention to the warnings that appear for a safe high reputation website which user visited before.

For creating less but effective warnings several suggestions were presented. Taking into account the design guidelines given by[15], can help to design an appropriate warning which will have a great impact on users decision. Creating active warnings was another suggestion studied by[2]. Active warnings showed to have a better effect compared to passive indicators since they interrupt the user's primary task and force her to notice the warning and take an action.

Beside focusing just on warning's presentation, other items such as social psychological factors like appeal to authority, social influence, concrete and vague threat can also have a great impact on user's behavior. Scammers used the mentioned items to fraud the victims by introducing themself as an authority figure and by gaining the users trust. The authors [3] mentioned it is best to create concrete warnings. This means that instead of using just a phrase like "'this site might harm your computer"', it is better to use phrases like "' This site wants to steal your bank account details"'. The user needs to get a clear illustration about the consequences of ignoring the warning.
Trusted authority figure was another factor presented in [3] which showed to have a great impact on users behavior. The individuals trusted the warnings that came from a trusted authority.

However, at the end every person has to decide whether she wants to pay attention to the alarms or not. Warning designers should be more accurate in creating indicators. Warnings should be more intelligent and should avoid interrupting the user with useless SSL warnings which are mostly false alarms,. On the other hand they should block the user's action when it is a serious dangerous situation . Beside that people need to increase their knowledge and have got to be carefully taught about security threats. The information about the existing frauds has to be spread very soon through the society.

5. REFERENCES

[1] Akhawe, D., Felt, A. P. : *Alice in Warningland: A Large-Scale Field Study of Browser Security Warning*

Effectiveness, Paper presented at the USENIX Security Symposium, Washington, D.C, 2013

[2] Egelman, S., Cranor, L. F., Hong, J: *You've Been Warned: An Empirical Study of the Effectiveness of Web Browser Phishing Warnings*, New York: Assoc Computing Machinery, 2008

[3] Modic, David and Anderson, Ross J: *Reading this May Harm Your Computer: The Psychology of Malware Warnings*, Available at SSRN: http://ssrn.com/abstract=2374379 or http://dx.doi.org/10.2139/ssrn.2374379,January 3, 2014

[4] Egelman, S., Schechter, S: *The Importance of Being Earnest [in Security Warnings]*,Paper presented at the Financial Cryptography and Data Security 2013, Okinawa, Japan.

[5] Lee, J. D., See, K. A: *Trust in automation: Designing for appropriate reliance. Human Factors*, 2004

[6] Krol, K., Moroz, M., Sasse, M. A: *Don't work. Can't work? Why it's time to rethink security warnings*, Paper presented at the Risk and Security of Internet and Systems (CRiSIS), 2012 7th International Conference on.

[7] Almuhimedi, Hazim, et al: *Your Reputation Precedes You: History, Reputation, and the Chrome Malware Warning.*, Symposium on Usable Privacy and Security (SOUPS).,2014.

[8] Herley, C: *So Long, And No Thanks for the Externalities: The Rational Rejection of Security Advice by Users.*, Paper presented at the NSPW '09 Proceedings of the 2009 workshop on New security paradigms Oxford, UK

[9] Murphy, K: *The Role of Trust in Nurturing Compliance: A Study of Accused Tax Avoiders*, Law and Human Behavior, 28(2), 187-209. 2004

[10] Kahan, D.M: *Social Influence, Social Meaning, and Deterrence*,Virginia Law Review, 83(2), 349-395 1997

[11] Modic, D., Lea, S. E. G : *Scam Compliance and the Psychology of Persuasion*, Journal of Applied Social Psychology. 2013

[12] Bikhchandani, S., Hirshleifer, D., Welch, I : *A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades*, The Journal of Political Economy, 100(5), 992-1026.1992

[13] Bearden, W.O., Netemeyer, R.G., Teel, J.E : *Measurement of Consumer Susceptibility to Interpersonal Influence*, Journal of Consumer Research, 15(4), 473- 481.1989

[14] *http://fraudavengers.org/scams/*

[15] Bauer, L., Bravo-Lillo, C., Cranor, L., Fragkaki, E. : *Warning Design Guidelines (C. S. Laboratory, Trans.)*, Pittsburgh, PA: Carnegie Mellon University.2013

[16] Titus, R. M., Dover, A. R : *Personal Fraud: The Victims and the Scams*, Crime Prevention Studies, 12, 133-151.2001

Hardware-accelerated Galois Field Arithmetic on the ARMv8 Architecture

Markus Ongyerth
Advisor: Stephan Günther
Seminar Future Internet WS1415
Chair for Network Architectures and Services
Department of Computer Sience, Technische Universität München
Email: ongyerth@in.tum.de

ABSTRACT

A limiting factor for throughput of a network is the limit of throughput achievable with traditional routing. Network coding is a way to avoid this problem. A limiting factor for network coding is its inherent arithmetic complexity. This is particular true for high-throughput networks, but lower throughput and embedded systems suffer from the same limitations. This paper evaluates the performance, of different implementation and algorithms doing the descrete math required for network coding on an ARMv8 architecture and compares it to on an ARMv7 architecture. Since the ARMv7 is a 32bit architecture while ARMv8 is a 64bit, this benchmark shows the advantage of having larger general purpose registers. The different implementations compared in this paper also show the performance gain by taking advantage of the NEON SIMD extensions, which increase register size (even more) to 128bit.

1. INTRODUCTION

In theory, network coding allows to increase the throughput of a network to its upper bounds [6]. It uses intelligent broad- and multicasting to distribute information in a way that is more efficient than routing. To achieve this, packets are aggregated and encoded in a way that allows a receiving node to decode the original packets. One of the easiest examples to show when and how network coding increases the throughput of a network is the butterfly network which is shown in Figure 1a.

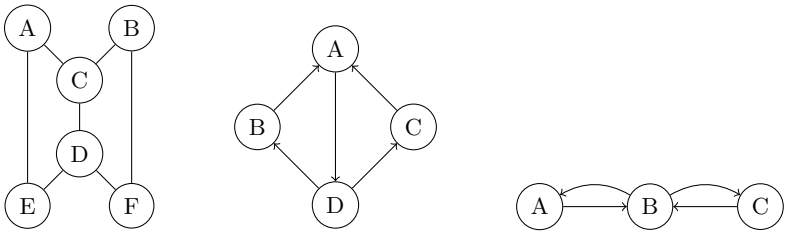
Each connection can transmit one packet of data per unit of time. The advantage of network coding in this kind of network becomes apparent when nodes A and B want to send data to nodes E and F. These nodes do not have to be the actual source or actual destination for this data but can be intermediate nodes as well. In a network that does not use any network coding this means that the connection between nodes C and D has to be used twice (once for A to F and once for B to E). If the network utilizes network coding, this can be avoided. In this case, node C receives both packets from nodes A and B separately and encode those into one common packet. Since there are only two packets to combine, this encoding can be done with a simple XOR operation. This common packet can than be transfered to node D as one packet. Node D then transmits the packet to nodes E and F and thus the connection between C and D is used only once. Nodes E and F receives the packet from A (for E) or the packet from B (for F) over their direct connection to the source node. With one of the source packets and the combined packet they are now able to decode the combined packet. Therefore obtaining both packets in a way that puts less strain on the transport medium.

In this example there was only one situation with only two packets that could benefit from network coding. For wireless networks with their inherent broadcast nature every packet two packets sent at the same time collide with each other, voiding both packets. This slows down the network since at a given time, only one, of possibly many, nodes in range of each other can send a packet without creating a collision. As the network gets more complicated–especially in mesh-networks–this becomes a problem. With network coding information can be exchanged over a network using less transmissions. By sending fewer packets the medium can be used more efficiently which increases the overall throughput of the network.

As mentioned, in this example the encoding and decoding of packets can be done with XOR operations. For more complicated networks, it is necessary to combine more than two packets and therefore more complicated encoding and decoding algorithms have to be applied. These algorithms have an inherent computational complexity, which prevents their practical deployment. In [4] Günther et al. have created and published a library, that does efficient finite field arithmetic needed for network coding. This library contains algorithms optimized to use SIMD extensions, in the case of ARM the NEON extension. They published benchmark results created with this library on an x86 CPU and ARMv7 CPU. The ARMv7 is a Cortex A15 and has a 32bit architecture. In this paper we compare the benchmark results of the library running on an ARMv8 processor, which has a 64bit architecture. This paper does not compare results to a x86 CPU, since traditionally x86 is aimed at high power and high performance while ARM is tuned for low power consumption. Therefore the ARM CPUs have a considerable lack of performance compared to the x86 CPU. At the time this paper is written the only processor comercially available with an ARMv8 architecture is the Apple A7, which is used for the benchmarks in this paper.

The remainder of this paper is organized as follows: first Section 2 describes Galois fields. Section 3 introduces and describes the hardware in this paper, while Section 4 introduces the algorithms used. In Section 5 the benchmark

doi: 10.2313/NET-2015-03-1_07



(a) Butterfly network (b) Diamond network (c) Wireless relay

results are presented and evaluated especially in comparison to the results in [4] made on the ARMv7 CPU. In Section 6 a few general remarks are made. Section 7 concludes the paper.

2. GALOIS FIELD

A Finite field, also called Galois field. We denote them as $\mathrm{GF}(p^n)$, where p is a prime number and n is a positive integer. This is possible, because the number of elements in a finite field is always a power of a prime number and all finite fields with the same size are isomorphic [4]. In this paper only binary extensions fields are considered, i.e., where the number of elements in the field is of order $q = 2^n$. Elements of this field can be expressed as polynomials over $\mathbb{F}_2$ of degree $n-1$, i.e.,

$$F_q[x] = \left\{ \sum_{i=0}^{n-1} a_i x^i \middle| a_i \in \mathbb{F}_2 \right\}. \quad (1)$$

The coefficients $a_i \in \mathbb{F}_2$ are represented by individual bits which allows for efficient processing on today's processor architectures. For the scope of this paper we focus on the finite fields of order $n = \{2, 4, 16, 256\}$, namely GF(2), $\mathrm{GF}(2^2)$, $\mathrm{GF}(2^4)$ and $\mathrm{GF}(2^8)$. These are the most important fields for network coding since the overhead for their coefficients is still kept relatively small compared to larger fields such as $\mathrm{GF}(2^{16})$ and $\mathrm{GF}(2^{32})$ and their elements naturally fit into processor registers.

Addition of $a, b \in F_q[x]$ is defined as:

$$a(x) + b(x) = \sum_{i=0}^{n-1} a_i x^i + \sum_{i=0}^{n-1} b_i x^i = \sum_{i=0}^{n-1} (a_i + b_i) x^i. \quad (2)$$

Note that coefficients are added according to the rules of the additive group associated with $\mathbb{F}_2$, meaning that addition is done modulo 2. Addition modulo 2 reduces to a simple XOR operation.

For Multiplication a polynomial r of degree n, that is irreducible over $F_q[x]$ is required. Irreducible means, that the polynomial r cannot be expressed as the product of two polynomials in $F_q[x]$. Such a polynomial is guaranteed to exist as shown in [5], but generaly not unique. Multiplication in the obtained finite field depends on the polynomial g. The product of $a, c \in F_q[x]$ is the unique remainder

$$b(x) = (a(x) \cdot c(x)) \bmod r(x). \quad (3)$$

The polynomial r is sometimes called *reduction polynomial* since it constrains the maximum degree of the result $b \in F_q[x]$. Because r is irreducible, it is guaranteed that $a(x) \cdot b(x)$ does not equal $r(x)$. This ensures that the reduction does not reduce a polinomial to zero and therefore that the multiplication result is–except for commutativity–unique.

Data words of n bit length are expressed as polynomials $a \in F_q[x]$. A vector $\underline{a} \in F_q^k[x]$ is used as representation of a data packet of length kn bit. A *generation* of N source packets can then be written as matrix $A = [\underline{a}_i \dots \underline{a}_N]$. A coded packet is obtained by

$$\underline{b} = Ac = \sum_{i=1}^{N} \underline{c}_i \underline{a}_i, \quad (4)$$

where $c^T = [c_1 \dots c_N] \in F_q^N[x]$ denotes a vector of random coefficients which are drawn independently uniformly distributed from $F_q[x]$.

3. THE HARDWARE USED

Feature	Apple A7	Exynos5
Frequency	1.4 GHz	1.4 GHz
Cores	2	4(+4)
L1 Cache	64 kB/64 kB	32 kB/32 kB
L2 Cache	1 MB/	2 MB

Table 1: The hardware specifications of the devices used. Apple specification extracted from [1] and [3], while the Exynos5 specifications are from [4]

The Apple A7. The device used for the benchmarks that are newly made for this paper is an Apple iPad Mini, second generation. This device is used because at the time this paper is written (September 2014) the Apple A7 is the only commercially available processor with an ARMv8 and therefore an 64bit ARM architecture. There are different devices that use an Apple A7 processor, but the frequency the processor runs on does not differ much between the devices (1.3 GHz to 1.4 GHz). The frequency on the iPad matches the frequency of the board used for comparison.

The problem imposed by using a device with an Apple A7 is that Apple has not published much information about the processor's specification or even its frequency. Fortunately, others are interested in the technical specification of these

doi: 10.2313/NET-2015-03-1_07

devices as well. *Anandtech* has published an article, about the iPhone 5s and later the iPad Mini, which both use the same Apple A7 SOC, analyzing the processor specification, of this platform. The information in 1 is extracted from those articles.

The Exynos 5 Octa. The device used for comparisons in this paper, is a ODROID-XU Lite development board. Table 1 displays the specification of this device.

The Exynos5 Octa follows the ARM "big.LITTLE" system and actually has 4 "big" cores and 4 "small" cores, but only one of these groups is active at a time. This benchmark was always executed on the faster cores. The core-count itself, however, does not really have an impact on the results since the library is single-threaded.

Looking at raw numbers, the two processors are similar for this benchmark. The A7 has twice as much L1 cache as the Exynos, but the Exynos has twice the L2 cache. The two big differences between the two platforms are the core count - 2 to 8 or rather 4 - and the register width of the processors. But as mentioned before, the advantage in core-count does not matter for this benchmark since it is aimed at single-core throughput, and the difference in word width is one of the main points why those platforms are compared to each other.

4. THE ALGORITHMS

For a base performance to compare against, a simple table lookup algorithm is used. For this table lookup all possible products of two elements of the Galois field are precomputed and saved in an array. The multiplication is then done by retrieving those values from the array.

The *imul* [4] algorithm does not require SIMD extensions and can therefore be implemented using only general purpose registers. It can also benefit from wider SIMD registers if available. It is suitable to run on microarchitectures that does not have SIMD extensions. The downside to this algorithm is that it scales badly with the degree of the finite field used. The library contains different implementations of this algorithm, which differ in the register size used. There is a 32 bit version and a 64 bit version. Both those versions only use general purpose registers. There is also a version using SIMD registers and instructions. Those registers are at least 64 bit–128 bit on both platform used in this paper.

The *shuffle* [4] algorithm benchmark results are not considered in this paper since the benchmark for the iPad has to be built, with the **Apple-llvm** compiler, which currently does not support the intrinsics used for this algorithm. (September 2014)

5. MEASUREMENTS

The linear encoding throughput is compared using a generation size of $N = 16$. The *throughput* is defined as the total size of encoded packets over a time interval measured in Gbit/s. The benchmark is done for packet sizes ranging from 128 B to 8 MB. As baseline performance the table lookup is used. The packet size has a significant impact on the performance since there is overhead that has to be done for every packet. This overhead amortizes for larger packets. Therefore, a larger packet size yields a higher throughput. This general assumption holds true until the memory requirements of the working set exceed the cache sizes.

When the cache sizes are reached, the throughput becomes limited by memory performance. Figures 1a to 1d show the performance of the Apple A7 compared to the Exynos5 in GF(2) and GF(2^2). Generally speaking, the A7 is about 2 to 3 times faster than the Exynos5. This rule of thumb is not true for every packet size and at its peak the Exynos5 is even faster than the A7. The change in throughput for varying packet sizes is rather similar for both processors used.

As expected, the throughput increases until the memory requirements of the working set hit the L1 cache limit. After this point the throughput remains about the same until the memory required for the working set hits the size of the L2 cache. At this point there is a second degradation in throughput, since main memory performance now has an effect on the algorithms. This degradation of throughput is most visible in the NEON implementation. The Exynos5s performance drops drastically to about half its value. The A7s memory does not throttle the performance as significant, but the impact is still visible. A bigger difference between the two platforms is visible when the difference between the 64 bit and the 32 bit on a single platform implementations is analyzed. On the Exynos5 the *imul 64bit* is at best as fast as the *imul 32bit* or even slower. This is caused by the lack of native 64bit operations, which cut the performance on 64bit numbers at least in half. The A7, which has support for native 64bit operations, shows that this assumption is correct. On the A7 , the throughput behaves comparable to the results [4] got on `x86_64`. Namely the *imul 64bit* algorithm is about 1.5 times to twice as fast as the *imul 32bit*. This performance increase is possible because the CPU is able to process twice the amount of data per instruction than it can process in a single instruction on 32bit registers. This advantage gained by using larger registers is visible by looking at the *imul NEON* implementation as well. This implementation uses NEON SIMD extensions and therefore has 128bit wide registers, which makes it another 1.5 to 2 times faster than *imul 64bit*.

Figures 1e to 1h show the benchmark results for GF(2^4) and GF(2^8). In these larger fields the throughput does not change as much with packet size. It seems like the throughput gets more consistent with the size of the finite field the arithmetic is done in. The A7 shows nearly no change when it hits the cache sizes. Because the throughput gets more limited by the CPU performance and less by memory performance. The impact on throughput by exceeding cache size gets less significant for larger field sizes on the Exynos5 as well even though, it isn't as good as on the A7. The difference between the two processors does not divert much from the observations made for smaller field sizes. The A7 is still about 2 to 3 times faster and benefits from 64bit general purpose registers when possible.

Next to the cache limits, another interesting packet size is 1024Kib, especially in comparison of the two processors. The A7 has an unusual big drop in throughput here while the Exynos5 has an irregular increase of throughput at this point. So far there is no conclusive explanation for this

doi: 10.2313/NET-2015-03-1_07

anomaly on either of the platforms.

6. REMARKS

[4] concluded, that the trend to heterogenous microarchitectures where both CPU and (integrated) GPU access the same memory might bring a similar gain when low level arithmetic is outsourced to the graphics processor.

It seems that there is an SRAM cache on the A7 SOC [2], that may be also accessible by the integrated GPU. For now this is guesswork and there is no easy way to prove and or test since Apple does not release any information about the SOC. But potentially this may be used to transfer data to and from the GPU and bypass the usual high overhead induced by memory transfer from CPU to GPU, which makes these operations on GPU, far more viable.

Something interesting to note about the two platforms compared in this paper is, although with different frequencies both of the platforms have an actual real live use case and those use cases are rather similar: they have been used as SOC for a smartphone. The Apple A7 is also used in the iPhone 5s and the Exynos5 is used in one version of the Samsung Galaxy S4. Both have been released in 2013 and competed on the market. Comparing the two platforms with this benchmark is not fair, since the benchmark only uses a single core and the Exynos5 which is significantly slower in this paper has more cores.

7. CONCLUSION

With scalar implementation on general purpose registers all field sizes larger than GF(2^2) are limited to less than 1 Gbit/s on the A7. Using SIMD extensions GF(2^4) reaches 1 Gbit/s. The performance of the *imul NEON* is about twice the performance of the *imul 64 bit*. The *shuffle* algorithm provided by *libmoepgf* cannot be used for a benchmark on the A7 yet. This algorithm outperforms the *imul* implementations on all tests performed [4]. It will be interesting to see whether or not the *shuffle* algorithm outperforms the algorithms on the A7 as well, and, if it does, what kind of performance it achieves on the A7.

References

[1] *Anandtech: The iPhone 5s Review*, `http://www.anandtech.com/show/7335/the-iphone-5s-review/3`

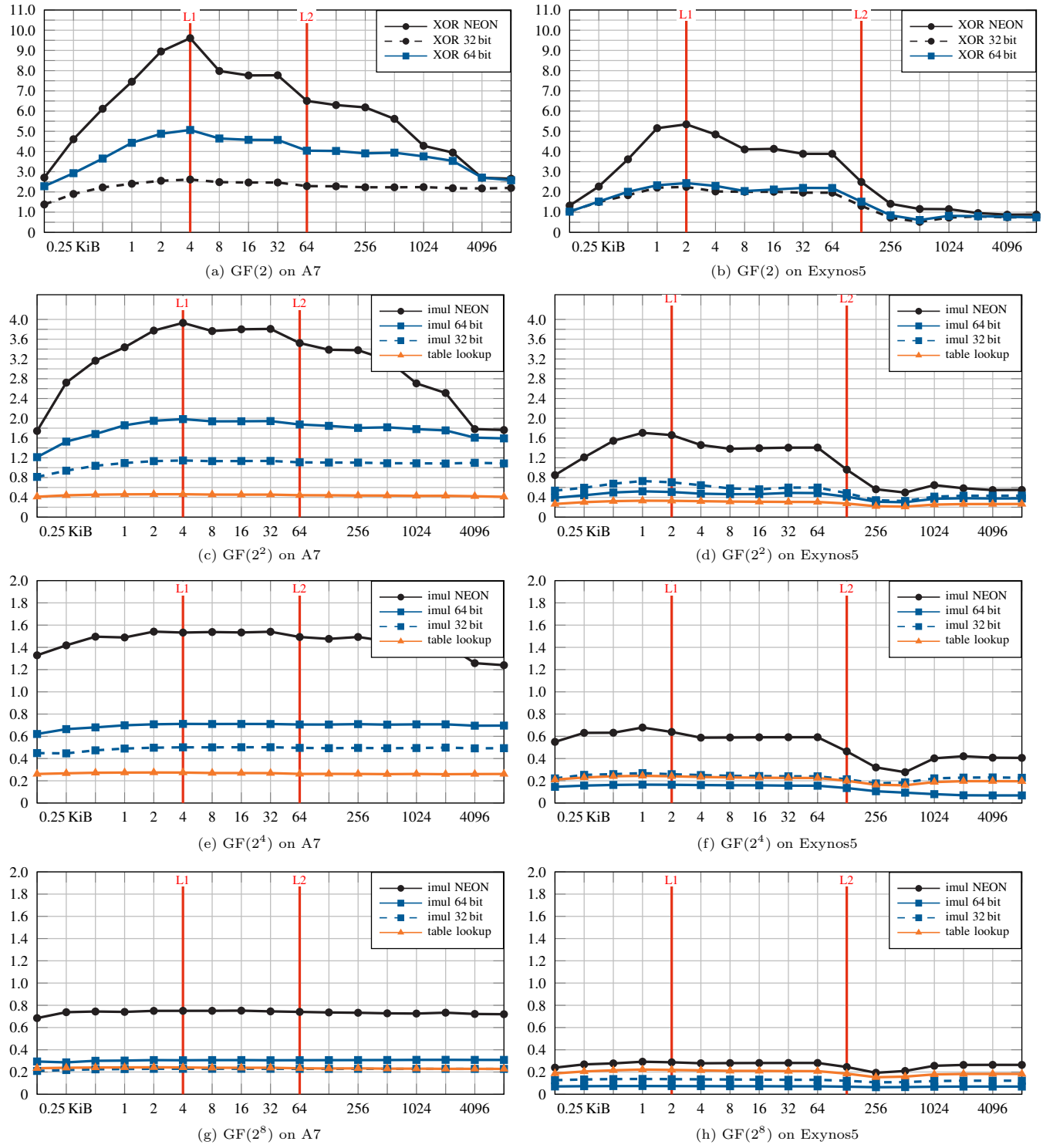
[2] *Anandtech: The iPad Air Review*, `http://anandtech.com/show/7460/apple-ipad-air-review/2`

[3] *Anandtech: The iPad Air Review*, `http://anandtech.com/show/7460/apple-ipad-air-review/3`

[4] Stephan M. Günther, Maximilian Riemensberger, Wolfgang Utschick: Efficient GF Arithmetic for Linear Network Coding using Hardware SIMD Extensions

[5] D. Hankerson, A. Menezes, and S. Vanstone: Guide to Elliptic Curve Cryptography, 1st ed., Jan. 2004.

[6] Shuo-Yen Robert Li, Senior Member, IEEE, Raymond W. Yeung, Fellow, IEEE, and Ning Cai: Linear Network Coding, IEEE transactions on infromation theory, vol. 49, no. 2, february 2003



(a) GF(2) on A7

(b) GF(2) on Exynos5

(c) GF(2^2) on A7

(d) GF(2^2) on Exynos5

(e) GF(2^4) on A7

(f) GF(2^4) on Exynos5

(g) GF(2^8) on A7

(h) GF(2^8) on Exynos5

Figure 1: Encoding throughput [Gbit/s] **for packet sizes varying from** 128 B **to** 8 MiB **in a generation of size 16 on Apple A7 (left) and Samsung Exynos5 Octa (right) both at** 1.4 GHz **with marks for both L1 and L2 cache sizes.**

doi: 10.2313/NET-2015-03-1_07

Measuring Privacy

Markus Schnappinger
Betreuer: Marcel von Maltitz
Seminar Future Internet „WS2014/15
Lehrstuhl Netzarchitekturen und Netzdienste
Fakultät für Informatik, Technische Universität München
Email: ga67wap@mytum.de

KURZFASSUNG

Ein Energiekontrollsystem in einem Gebäude sammelt kontinuierlich Sensordaten. Bei der Abdeckung verschiedener Anwendungsfälle muss jedoch eine Privatsphäre-wahrende Darstellung dieser teils sensiblen Informationen gefunden werden. Nach einer Untersuchung von etablierten Methoden aus anderen Fachbereichen werden Möglichkeiten aufgezeigt in den einzelnen Szenarien einen geeigneten Datenschutz zu gewährleisten. Durch Anwendung der empfohlenen Verfahren ist es möglich, jeder Rolle Zugang zu von ihr benötigten Informationen zu geben ohne eine Überwachung der Nutzer oder das Erstellen eines Verhaltensprofils fürchten zu müssen.

Schlüsselworte

Privacy, Data Mining, Energiekontrollsysteme, IDEM

1. EINLEITUNG

Mit einer zunehmend energiebewussteren Lebensweise vieler Menschen gewinnen auch Energiekontrollsysteme in Gebäuden an Bedeutung. Derartige Systeme ermöglichen eine sekundengenaue Überwachung aller Stromflüsse durch Sensoren an den Verbrauchern und Leitungen. Neben Privathaushalten, die sich einen genauen Überblick über ihren Energieverbrauch verschaffen wollen, sollen derartige Systeme auch in öffentlichen Einrichtungen und Unternehmen Verwendung finden, um auch dort die Nutzer für Energieeinsparungen zu sensibilisieren und ein mögliches Einsparpotenzial aufzuzeigen. Um genaue zeitliche und lokale Informationen bereitstellen zu können, muss ein solches System mithilfe einer Vielzahl von Sensoren permanent hochauflösend Daten sammeln und zum späteren Abruf verwalten. Die so generierten Energiemessdaten können anschließend von verschiedenen Benutzergruppen in unterschiedlicher Granularität abgerufen werden; beispielsweise sollte jeder Nutzer den durch ihn verursachten Energieverbrauch mit hohem Detailgrad einsehen können, während ein Buchhalter hingegen zu Abrechnungszwecken lediglich den Gesamtverbrauch einer räumlichen Komponente über einen längeren Zeitraum benötigt. Gemäß des Datensparsamkeitprinzips „so wenig Daten wie möglich“ ist es notwendig dem Buchhalter in diesem Szenario weniger genaue Informationen bereitzustellen als dem Mitarbeiter selbst. So erhält jede Rolle Zugriff auf gerade so viele Informationen in gerade so genauer Auflösung wie sie zur Erfüllung der ihr zugedachten Aufgabe benötigt. Auf diese Weise wird ein Missbrauch der Daten beispielsweise zur Überwachung der Mitarbeiter anhand ihrer gesammelten energetischen Datensätze erschwert.

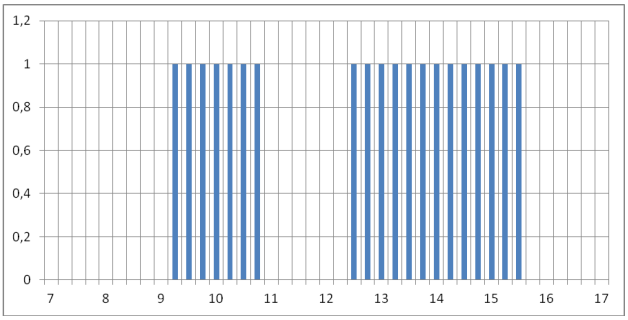


Figure 1. Beispielgraph eines Sensors

Zur Verdeutlichung dieser Gefahr nehmen wir an, der obige Verlauf (Figure 1) zeige die Messwerte eines Sensors, der den Stromverbrauch der Beleuchtung eines Ein-Mann-Büros aufzeichnet. So lässt sich aus diesen Werten folgern, der Mitarbeiter hat nicht um 9 Uhr sein Büro betreten, war nicht bis 17 Uhr im Raum und hat außerdem um die Mittagszeit seinen Arbeitsplatz für 75 Minuten verlassen. Bei Arbeitszeiten von 9 bis 17 Uhr und einer 45-minütigen Mittagspause würde dieser Datensatz vermutlich nicht zu seinen Gunsten interpretiert werden.

Entwickelt wird nun ein System, welches die gesammelten Daten sicher und sinnvoll verarbeitet und dabei die Einhaltung von Privatheits- und Datenschutzkriterien geeignet gewährleistet. Nach einem Versuch den Privatheitsbegriff intuitiv zu erfassen werden im Folgenden unterschiedliche Methoden zur Handhabung sensibler Daten, teils aus anderen Fachbereichen, vorgestellt und hinsichtlich ihrer Anwendbarkeit auf Energiekontrollsysteme in Gebäuden untersucht. Daran anschließend wird versucht diese hinsichtlich der hier gegebenen Anwendungsdomäne zu adaptieren und eigene Lösungsansätze sowie der aktuelle Forschungsstand werden präsentiert. Als besonderer Anwendungsfall wird das Wettkampfspiel EQ eingeführt, welches einen zum Energiesparen motivierenden Vergleich der Mitarbeiter bereitstellt und dabei keine Privatheitskriterien verletzt. Abschließend wird evaluiert, dass kein Ansatz alle Anwendungsfälle zufriedenstellend bedient und die Verwendung eines Methodensets die beste Option darstellt.

2. Intuitiv-naiver Privatheitsbegriff

In diesem Kapitel wird versucht den Begriff Privatheit zunächst informell zu erfassen und zu verdeutlichen welche Daten als schützenswert erachtet werden sollten. Nötig ist der Umgang mit sensiblen Informationen in unserem Alltag häufiger als man

doi: 10.2313/NET-2015-03-1_08

zunächst annehmen mag. In der Apotheke, im Reisebüro, bei der Bank, der Passkontrolle am Flughafen, beim Arzt, und an vielen weiteren Orten finden sich Abgrenzungen und Hinweisschilder, die zum Einhalten eines gewissen Abstands zur Sicherung der Privatsphäre ermahnen. Niemandem ist wohl dabei wenn ein anderer, womöglich nicht vertrauenswürdiger Mitmensch bestimmte Informationen über uns erlangt oder erlangen könnte. Meist lassen sich diese den Aggregationen Medizin und Finanzen zuordnen, sowie Informationen, welche zur Erstellung eines Verhaltensmusters oder Persönlichkeitsprofils verwendet werden könnten. Psychologisch lässt sich dies erklären durch die Angst, diese Daten könnten zu unserem Nachteil interpretiert werden. Eine bekannte Krankheit lässt einen Menschen schwach wirken, ein niedriges Gehalt weckt Schamgefühl, ein hohes Verlegenheit.

Barrieren beim Arzt oder bei Kreditinstituten erfüllen also offensichtlich direkt unser Bedürfnis, diese Daten zu schützen. Doch warum finden sie sich auch, zum Beispiel, in Apotheken? Dort wird lediglich über Medikationen gesprochen, nicht jedoch über Krankheiten und Gebrechen per se. Dennoch ist der Schutz der Daten auch hier sinnvoll. Die Argumentation bedarf dabei nur eines einzelnen, aber wesentlichen Schrittes, nämlich des Rückschlusses von einem ausgegebenen Medikament zum damit behandelten Gebrechen. Treffe ich meinen Nachbarn in der Apotheke beim Erstehen einer Hämorrhoidensalbe an, beeinflusst dies unser Verhältnis in gleichem Maße als wäre ich selbst bei der Diagnosestellung anwesend gewesen. Schützenswert sind also nicht nur Informationen, die wir geheim zu halten wünschen, sondern auch alle Daten, die direkt oder im gegenseitigen Zusammenwirken Rückschlüsse darauf erlauben. Im Umkehrschluss bedeutet dies auch, dass beim Umgang mit sensiblen Daten, zum Beispiel in einem Krankenhaus, sichergestellt werden muss, dass die betreffende Person nicht identifiziert werden kann. Um dies zu verhindern findet sich in der Medizin das Safe-Harbor-Prinzip, welches einen Katalog an sogenannten Identifiern bereitstellt, die einen Rückschluss auf eine Patientenidentität ermöglichen könnten und deshalb nicht veröffentlicht werden dürfen. Neben Namen, Telefon-, Fax- und Konto- sowie Sozialversicherungsnummern, gehören auch URLs und IP-Adressen ebenso dazu wie Photographien des Gesichts und vergleichbare Bildnisse, Fingerabdrücke oder Stimmaufzeichnungen, Autokennzeichen und Seriennummern, Seriennummern möglicher Implantate oder der Krankenhausdokumentation sowie alle temporalen Angaben über Geburt (ausgenommen das Jahr), Aufnahme, Entlassung, möglicherweise Tod. Sollte ein Patient über 89 Jahre alt sein und kein Zusammenfassen mit weiteren Personen dieser Altersgruppe möglich sein, müssen alle Hinweise auf das Alter gänzlich gestrichen werden. Zusätzlich dürfen geographische Angaben, die detaillierter als ein Staatsgebiet sind, nur in Form der ersten drei Ziffern des Postleitzahlgebietes aufgeführt werden. Sollten in einem Gebiet, welches durch diese drei Ziffern beschrieben wird, weniger als 20.000 Menschen wohnhaft sein, wird die Angabe durch 000 ersetzt. [1, 4] Während die erstgenannten Attribute intuitiv einleuchten, wirkt vor allem das letztgenannte ungemein kompliziert. Erwähnt sei an dieser Stelle, dass obiger Katalog in den USA entwickelt wurde. Studien dort zeigten, 87 Prozent aller Amerikaner können durch den Verbund aus fünftstelliger Postleitzahl, des Geschlechts und des Geburtsdatum identifiziert werden. [2] Obige Restriktion soll also einen solchen Schluss verhindern. Im Folgenden unterteilen wir Attribute analog zu [3] in drei Gruppen: Daten, die direkt auf eine Person schließen lassen, nennen wir explizite Identifier; einen Informationsverbund wie Geburtsdatum, Geschlecht und Postleitzahl, der im Zusammenspiel eine Identifikation ermöglichen könnte, bezeichnen wir als Quasiidentifier; schützenswerte Daten betiteln wir als sensibel.

Zunächst stellt sich nun die Frage, warum man nicht auch Quasiidentifier gemäß des Safe-Harbor-Prinzips verheimlicht. Die Antwort liegt in der verschwindend geringen Verwertbarkeit eines so modifizierten Datensatzes. Wir wählen als Beispiel eine klinische Studie über das Auftreten einer neuartigen Krankheit, bei der sowohl alle Identifier als auch Bestandteile des Quasiidentifiers unterdrückt werden. Zurück bleibt lediglich eine Liste von positiven und negativen Testergebnissen, aus denen sich keine räumliche Häufung, erhöhte Infektanfälligkeit einer bestimmten Altersgruppe, eines Geschlechts oder ähnliche wissenschaftlich interessante Eigenschaften manifestieren. Analog verhält es sich bei einer Psyeudonymisierung der Quasiidentifier, also dem Ersetzen der Daten durch beispielsweise Zufallszahlen oder Angaben ohne Bedeutung. Der Informationsgehalt ist identisch. Bei Fragen der nationalen Sicherheit mag dies unter Umständen eine praktikable Lösung sein, doch wenn wissenschaftlich verwertbare Informationen erhalten bleiben sollen, muss eine andere geeignete Sicherstellung der Privatheit gefunden werden. [2]

3. Gängige Praxis

Betrachten wir nun Maßnahmen, wie sie üblicherweise zum Schutz von vertraulichen Daten getroffen werden. Wir gehen dabei davon aus, dass die Originale eines Datensatzes unter Verschluss gehalten werden und lediglich Modifikationen desselben veröffentlicht werden. Um unbefugten Zugriff auf die Originale zu unterbinden, müssen nach Bundesdatenschutzgesetz [6] § 9(Anlage) folgende Maßnahmen ergriffen werden: Zunächst ist eine Zutrittskontrolle zu installieren, das heißt der Zutritt zu den Verarbeitungsanlagen muss verwehrt werden – beispielsweise durch eine verschlossene Tür. Eine Zugangskontrolle verhindert zusätzlich unbefugten Systemzugang, etwa durch eine Passwortabfrage. Eine anschließende Zugriffskontrolle gewährleistet Zugriff auf die Daten ausschließlich gemäß einer bestimmten Rolle – beispielsweise nur lesend. Des Weiteren müssen Weitergabe-, Eingabe-, Auftrags- und Verfügbarkeitskontrollen installiert werden. Abschließend sei das unveräußerliche Recht eines Betroffenen auf Auskunft (§19,34) über erhobene Daten sowie auf Berichtigung, Sperrung und Löschung (§20,35) erwähnt. [4,6] Gesetzt den Fall der Schutz der vertraulichen Daten ist gegeben, widmen wir uns nun der Veröffentlichung potenziell sensibler Informationen.

3.1 k-Anonymität

Wie bereits erwähnt stellt das Unterdrücken oder Pseudonymisieren der Quasiidentifier aufgrund des zu hohen Informationsverlustes keine praktikable Lösung dar. Eine oftmals gebrauchte Methode eine Privatheit gemäß der in Kapitel 2 beschriebenen Kriterien zu gewährleisten besteht nun darin, lediglich Teile eines Attributs zu unterdrücken, sodass mehrere nicht unterscheidbare Tupel entstehen. Aus einer Postleitzahl 12345 wird etwa 123*; und Herr Schmidt aus 12345 ist nicht

länger von Herrn Schmidt aus 12346 und Herrn Schmidt aus 12347 unterscheidbar (Table 1).

Table 1. Beispieltabelle original (oben) und 3-anonymisiert

Name	Geschlecht	Postleitzahl
Schmidt	M	12345
Schmidt	M	12346
Schmidt	M	12347

Name	Geschlecht	Postleitzahl
Schmidt	M	1234*
Schmidt	M	1234*
Schmidt	M	1234*

Erhält man nun Kategorien mit jeweils mindestens k Tupeln, spricht man von k-Anonymität. Dies bedeutet jedes Tupel ist von mindestens k – 1 anderen nicht unterscheidbar. [2] Auf diese Weise wird die in Kapitel 2 formulierte Anforderung erfüllt, dass ein Individuum nicht eindeutig mit einem sensiblen Datum in Verbindung gebracht werden kann.

Im Weiteren verfeinern wir den durch diese Technik formalisierten Privatheitsbegriff anhand folgender Beispieltabelle, wie sie etwa in einem Krankenhausinformationssystem vorliegen könnte (Table 2).

Table 2. Krankenhaustabelle mit sensiblen Daten

	PLZ	Alter	Erkrankung
1	12344	21	Syphilis
2	12345	26	Filzläuse
3	12345	24	Chlamydien
4	12356	28	Gonorrhoe
5	12346	33	Beinfraktur
6	12348	44	Bandscheibenvorfall
7	12344	39	Krebs
8	12344	37	Krebs
9	12353	65	Herzinfarkt
10	12344	54	Herzinfarkt
11	12347	47	Herzinfarkt
12	12354	73	Herzinfarkt

Wir erreichen 4- Anonymität durch folgende Generalisierung (Table 3).

Table 3. 4-anonymisierte Krankenhaustabelle

	PLZ	Alter	Erkrankung
1	123**	<29	Syphilis
2	123**	<29	Filzläuse
3	123**	<29	Chlamydien
4	123**	<29	Gonorrhoe
5	1234*	30-45	Beinfraktur
6	1234*	40-45	Bandscheibenvorfall
7	1234*	30-45	Krebs
8	1234*	30-45	Krebs
9	123**	>=45	Herzinfarkt
10	123**	>=45	Herzinfarkt
11	123**	>=45	Herzinfarkt
12	123**	>=45	Herzinfarkt

Zweifelsfrei erfüllt diese Tabelle die Anforderung, dass kein Patient eindeutig identifiziert werden kann. Dennoch weisen solche k-anonymisierten Veröffentlichungen Schwächen auf; diese zeigen wir nun anhand eines Beispielszenarios auf. Angenommen, Anna arbeitet im Krankenhaus und hat Zugriff auf die obige anonymisierte Tabelle. Annas Mutter ist besorgt um ihren Nachbarn, der mit einem Krankenwagen in jenes Krankenhaus gebracht wurde, in dem Anna arbeitet. Ihre Mutter bittet also Anna nachzusehen, was ihrem Nachbarn fehlen könnte. Sein Alter schätzt sie auf 50 bis 60 Jahre. Obwohl Anna nicht eindeutig bestimmen kann, welches Tupel dem Nachbarn ihrer Mutter zugehörig ist, identifiziert sie dennoch einen Herzinfarkt als seine Erkrankung. Diese Schlussfolgerung beruht auf dem sogenannten Homogenitätsangriff nach [5]. Ein weiterer Angriff beruht auf vorhandenem Hintergrundwissen. Angenommen Anna sieht bei ihrer Tätigkeit im Krankenhaus einen Bekannten aus ihrem Sportverein zügig in ein Behandlungszimmer gehen. Da sie sein Alter von 37 Jahren kennt, schließt sie auf einen der Einträge 5-8. Allerdings kann sie die Einträge 5 und 6 mit dem zusätzlichen Wissen ausschließen, welches sie erlangt hat, als sie ihn selbstständig zu Fuß gehen sah. Da die Erkrankung der Einträge 7 und 8 identisch ist, schließt sie auf Krebs als Behandlungsgrund. Das Ausschließen von Einträgen aus einer Tupelkategorie kann also wieder zu einem Homogenitätsangriff führen [5]. Eine weitere Schwäche der k-Anonymität findet sich, wenn sich die Tupel sowohl in der Veröffentlichung als auch im originalen Datensatz in der gleichen Reihenfolge befinden und dem Angreifer das Sortierungsargument bekannt ist. Wurde zum Beispiel aufsteigend nach dem Alter sortiert (hier nicht der Fall), ist es wahrscheinlich, dass einem 45-jährigen Mann der erste Eintrag der Kategorie „>=45 Jahre alt“ zugehörig ist. Dieser Angriff kann durch Randomisieren unterbunden werden [2]. Eine weitere Angriffsmöglichkeit bieten mehrfache Veröffentlichungen, bei denen jeweils unterschiedliche Bestandteile des Quasiidentifiers aufgeführt werden. Man nehme zur Verdeutlichung eine Tabelle mit PLZ, Geschlecht und Krankheit sowie eine weitere mit den Einträgen Geburtsdatum und Krankheit. Ein Join der Tabellen über das gemeinsame Attribut Krankheit führt zum Quasiidentifier PLZ, Geschlecht und Geburtsdatum, der wie wir wissen ausreicht um einen Großteil aller Amerikaner zu identifizieren. Während die letzten beschriebenen Angriffe mit geringem Aufwand abgewehrt werden können, benötigt man für Hintergrundwissen-basierte und Homogenitätsangriffe eine Technik, die strenge Privatheitsanforderungen erfüllt als k-Anonymität.

3.2 l-Vielfalt

Eine Verschärfung der k-Anonymität bietet die l-Vielfalt. Wie oben dargelegt basiert ein Homogenitätsangriff auf der Tatsache, dass k-Anonymität zwar k nicht unterscheidbare Tupel bereitstellt, aber nicht gewährleistet, dass die mit ihnen verbundenen sensiblen Daten nicht identisch sind. Ziel ist es nun, dass ein Angreifer unabhängig von seinem Hintergrundwissen durch die anonymisierte Tabelle keinen Wissenszuwachs erlangen kann. Eine zu geringe Vielfalt in einem Block aus k Einträgen ermöglicht allerdings einen solchen Informationsgewinn und sollte daher vermieden werden. Deshalb wird die Schranke $l >= 2$, $l <= k$ eingeführt, welche aussagt, dass die l häufigsten Werte in jedem solchen Block ungefähr gleich häufig auftreten [5]. Auf diese Weise wird eine gewisse Heterogenität in den durch k-Anonymität generierten Blöcken erzeugt. Neben der Blockhomogenität kann wie erwähnt auch ein vorhandenes Hintergrundwissen eines Angreifers zu einem positiven Rückschluss führen. Um bei einer l-vielfältigen Tabelle einen Schluss ziehen zu können, benötigt man allerdings l-1 zusätzliche Informationen, da l-1 unzutreffende Tupel ausgeschlossen werden müssen. Auf diese Weise ist eine Parametrisierbarkeit der Metrik gegeben und die Strenge des Datenschutzes kann je nach Anwendungsfall angepasst werden. Somit stellt diese Technik eine nützliche Erweiterung der k-Anonymität dar, die den Homogenitätsangriff abwehren kann und Angriffe mit zusätzlichem Hintergrundwissen erschwert.

Problematisch ist dabei die Ungewissheit welcher Art ein potenzielles Hintergrundwissen ist [5]. Auch können bestimmte Einträge aufgrund von bekannten globalen Verteilungen, ethnischen Dispositionen etc. als wahrscheinlicher oder unwahrscheinlicher evaluiert werden [3]. Wir betrachten nochmals obiges Beispiel, als Anna ihren Sportkameraden im Krankenhaus identifizierte. Zwar waren hier bei vier zutreffenden Quasiidentifiern drei unterschiedliche Erkrankungen gelistet und damit eine ausreichende Vielfalt gegeben, dennoch konnte Anna zwei davon ausschließen und somit ein Krebsleiden folgern. Durch Anpassen der k- und l- Schranke durch den Veröffentlichenden hätte ein solcher Schluss erschwert werden können.

Neben menschlichen Faktoren offenbart folgendes Szenario eine weitere, schwerwiegendere Schwäche der l-Vielfalt. Ein weiteres Mal betrachten wir Anna, die wie gehabt Zugang zur Beispieltabelle *Table 3* hat. In einer Bar lernt Anna den Studenten Bernd kennen, den sie äußerst sympathisch findet. Im Laufe des Abends erwähnt Bernd er sei zum Zeitpunkt, der durch die Tabelle abgebildet wird, in Annas Krankenhaus Patient gewesen. Verunsichert durch diesen Umstand stellt Anna Nachforschungen an. Durch Bernds Äußeres und seinen Studentenstatus vermutet sie ein Alter unter 29 Jahren und fokussiert sich deshalb auf den ersten Block der Tabelle. Dieser ist 4-anonymisiert und aufgrund vier unterschiedlicher sensibler Einträge auch 4-vielfältig. Trotz dieser formal betrachtet exzellenten Privatheitskriterien fasst Anna bestürzt den Entschluss Bernd nicht wiedersehen zu wollen. Durchaus nachvollziehbar, offenbart doch die Tabelle trotz aller erfüllter Kriterien Bernds Infektion mit einer ansteckenden sexuell übertragbaren Erkrankung. Angesicht dieser Überkategorie ist es für Anna auch nur noch zweitrangig welche Infektionsart genau vorliegt.

Um auch auf Überkategorien basierende Schlüsse zu unterbinden, muss diese Formalisierung nochmals adaptiert und erweitert werden.

3.3 t-Nähe

Eine weitere Metrik namens t-Nähe beruht auf dem Ansatz die Verteilung eines sensiblen Datums in einem Äquivalenzblock möglichst der Verteilung dieses Datums im Gesamtvorkommen anzunähern. Bei erfüllter l-Vielfalt liegt die Wahrscheinlichkeit für eine richtige Zuordnung eines sensiblen Datums bei ca. $1/l$. Angenommen, eine Tabelle zeigt die Testergebnisse bezüglich einer negativ konnotierten Krankheit –beispielsweise AIDS. Ein Block in dieser k-anonymisierten Tabelle sollte entweder ausschließlich negative Testresultate beinhalten, da nicht krank zu sein keinen Nachteil mit sich bringt; oder gemäß oben 2-vielfältig sein, da mit positivem und negativem Testergebnis zwei Werte für das vertrauliche Attribut möglich sind. Folglich ergibt sich die Wahrscheinlichkeit $1/l$ beziehungsweise 0,5 dafür dass ein Mensch AIDS hat, dessen Eintrag in einem solchen Block vorkommt. Dies verletzt die Privatsphäre jenes Individuums, da diese Wahrscheinlichkeit sehr viel größer ist als die tatsächliche Wahrscheinlichkeit erkrankt zu sein, die in diesem Fall durch die Verteilung über alle Blöcke der Tabelle wiedergespiegelt wird. Dieses Problem versucht die t-Nähe zu bewältigen. Um dieses Ziel zu erreichen muss der Abstand zwischen zwei Attributen quantifiziert werden, was sich je nach Art der Attribute unterschiedlich gestaltet. Der kleinste zulässige Abstand einer Verteilung ist durch die Schranke t gegeben. Werden nominal skalierte Eintragungen untersucht, die keiner hierarchischen Ordnung folgen, beträgt dieser Abstand normiert 1. Alle möglichen Werte sind gleich unterschiedlich. Der Abstand ordinal skalierter Werte wie etwa numerischer Attribute, Skalen und ähnliches wird durch die Anzahl der zwischen ihnen liegenden Werte gekennzeichnet. Die Handhabung hierarchisch organisierter Nominalwerte hingegen erfordert Kenntnis dieser Hierarchie. Diese wird als Baumstruktur dargestellt; Blätter bilden die genauen Tabelleneinträge ab, Knoten mögliche Überkategorien. Alle Blätter besitzen zudem identische Tiefe n, welche den Nenner des Abstandes zweier Elemente bildet. Der Zähler stellt die Anzahl der Schritte dar, die man in der Hierarchie Richtung Wurzel unternehmen muss um einen gemeinsamen Vaterknoten beider Elemente zu finden [3]. Die oben verwendeten Krankheiten lassen sich wie folgt darstellen (Figure 2):

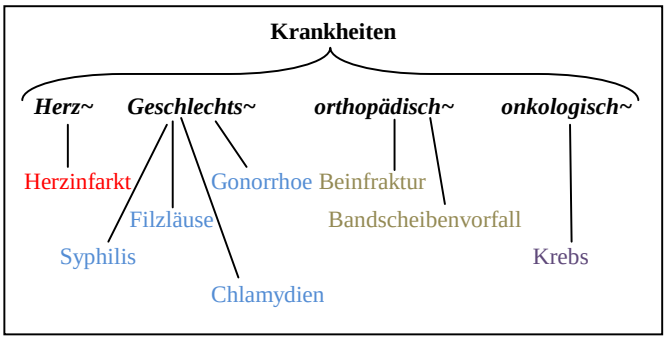


Figure 2. mögliche Hierarchie der dargestellten Krankheiten.

So beträgt der Abstand zwischen ‚Herzinfarkt' und ‚Krebs' 2, da erst in ‚Krankheiten' ein gemeinsamer Vorgänger gefunden wird. Zwischen ‚Beinfraktur' und ‚Bandscheibenvorfall' hingegen

doi: 10.2313/NET-2015-03-1_08

beträgt die Distanz 1, da beide den orthopädischen Gebrechen zuzuordnen sind.

3.4 Evaluierung der vorgestellten Methoden

Zwar lassen sich die vorgestellten Methoden zur Sicherung der Privatheit sensibler Daten leicht ineinander überführen und daher auch mit Hilfe von Adaptionen gängiger Algorithmen implementieren [5], so leidet dennoch mit zunehmender Sicherung der Daten auch deren Nutzbarkeit und wissenschaftliche Verwendbarkeit [3,5]. Begründet ist dies darin, dass bestimmte Einträge nur dann mit der gewünschten Technik aggregiert werden können wenn womöglich interessante Attribute oder Attributteile unterdrückt werden. Auch kann beobachtet werden, dass Äquivalenzblöcke einer anonymisierten Tabelle an Umfang zunehmen je mehr Kriterien zur Anwendung kommen [5]. Auch dies senkt nochmals die Verwendbarkeit einer Tabelle. Zusätzlich muss bei Verwendung der strengsten Richtlinie t-Nähe eine Hierarchie bekannt sein oder gegebenenfalls erstellt werden. Es muss nun evaluiert werden inwiefern die vorgestellten Metriken auf den eingangs erläuterten Kontext eines Energiekontrollsystems eines Gebäudes angewendet oder angepasst werden können.

4. Transfer zur Problemstellung

Während die in 3 dargestellten Techniken einen Schutz von sensiblen Absolutwerten in Tabellen bereitstellen können, sind bei einem Energiekontrollsystem zusätzlich zeitliche Verläufe schützenswert, da auch diese Dimension Aufschluss über Nutzerverhalten gibt, somit eine Möglichkeit zur Überwachung besteht und Profile abgeleitet werden können. Um dies zu unterbinden müssen die meist als Zeitreihen vorliegenden Informationen geeignet aggregiert werden. Die vorgestellten Techniken, wie sie zum Beispiel für medizinische Daten sinnvoll einsetzbar sind, können für diesen Anwendungsfall folglich nicht ohne Adaption übernommen werden, da die Daten in abweichendem Format vorliegen. Bevor in Kapitel 5 derartige Anpassungen erläutert werden, befasst sich dieser Abschnitt mit Problemen beim Transfer zur gegebenen Problemstellung und dortigen Anforderungen. Wie erwähnt existiert keine Isomorphie zwischen den tabellenbasierten Ansätzen aus Kapitel 3 und den hier relevanten zeitlichen Zusammenhängen, da von anderen Anwendungsfällen ausgegangen wird.

Als schützenswerte Eigenschaften eines Verlaufes können maximale und minimale Amplitude identifiziert werden sowie der Durchschnittswert während eines Zeitabschnittes. Gerade für Verhaltensprofile bedeutsam sind zudem markante Peaks zu bestimmten Zeitpunkten, ebenso wie erkennbare Trends und Entwicklungen. Privatsphäre-Anforderungen in diesem System beziehen sich demnach vor allem auf den Schutz dieser Daten in den gegebenen Anwendungsfällen. Ein solcher besteht beispielsweise darin, dass Nutzer die mit ihrer Person gekoppelten Daten unverfälscht und in höchster Detailstufe einsehen können. Während man auf die Energieinformationen anderer nur mit deren Einverständnis lesend zugreifen kann, ist eine zusätzliche Rolle des Energiemanagers zudem befugt diese von allen Mitarbeitern einzusehen. Buchhaltern hingegen stehen gemäß des Datensparsamkeitsprinzips nur stark aggregierte Werte und Summen zur Verfügung. Durch Public Displays, also Veröffentlichungen von Fakten in mittlerer Granularität, soll zudem eine Steigerung der Gebäudetransparenz möglich sein. Da folglich verschiedene Feinheitsgrade für verschiedene Szenarien erreicht werden müssen, muss der verwendete Privatheitsbegriff quantifizierbar und die entsprechende Technik parametrisierbar sein.

4.1 Motivationsmetrik EQ

Zur Verdeutlichung dass die Definition einer anwendungsspezifischen Metrik möglich ist, welche privatsphäreschützend ist und trotzdem einen Nutzen bietet, wird in diesem Abschnitt eine solche entwickelt.

Um bei den Nutzern des Gebäudes ein erhöhtes Bewusstsein für den individuellen Energieverbrauch zu schaffen ist ein spielerischer Wettkampf ein guter Antrieb. Dieser sollte durch einen Vergleich mit anderen einen Anreiz bereitstellen sich zu verbessern, ohne dabei sensible Daten und konkrete Werte zu offenbaren. Eine solche Privatsphäre-erhaltende Metrik namens EQ wird im Folgenden präsentiert. Inspiriert vom Konzept des Intelligenzquotienten, welcher die eigene Leistung mit dem durchschnittlich erzielten Erfolg in Relation setzt, wird auch hier ein solcher Ansatz verfolgt. Anstatt den Quotient aus dem Energieverbrauch des Nutzers und dem durchschnittlichen Energieverbrauch zu berechnen, betrachten wir hier allerdings den Kehrbruch desselben, da ein höherer Verbrauch zu einem niedrigerem Ergebnis führen sollte. Der EQ eines Nutzers berechnet sich demnach als $\frac{Mittelwert\ aller}{eigener\ Wert} \cdot 100$. Obwohl dies simpel erscheinen mag, erfüllt diese Metrik dennoch alle Anforderungen. Basierend auf seinem persönlichen Wert kann ein Teilnehmer lediglich auf den Durchschnittswert zurückschließen – in dessen Kenntnis kann er allerdings auch durch Public Displays gelangen. Außerdem beinhaltet diese Methode zwei psychologische Tricks: Zum Einen stimuliert es den menschlichen Trieb besser als der Durchschnitt sein zu wollen, zum Anderen wird durch die Multiplikation mit 100 auch ein optischer Anreiz gegeben einen dreistelligen Wert zu erreichen. Um den Zweck der Mitarbeitermotivation zu erfüllen reicht der Vergleich jenes mit dem Mittelwert aus, sich analog zum Intelligenzquotienten um eine Normalverteilung der Werte zu bemühen ist weder sinnvoll noch gerechtfertigt.

5. Adaption und neue Lösungsansätze

Dieses Kapitel beschäftigt sich mit Möglichkeiten die identifizierten Anwendungsfälle abzudecken und zeitgleich sensible Daten zu schützen. Nach einem Versuch die durch Graphen dargestellten Verläufe sinnvoll und Privatsphäre-erhaltend zu modifizieren wird der Ansatz verfolgt, diese Verläufe auf Tabellen abzubilden, sodass die Techniken k-Anonymität, l-Vielfalt und t-Nähe zur Anwendung kommen können. Anschließend werden Möglichkeiten zur tatsächlichen Aggregation von Zeitserien präsentiert sowie deren Anwendbarkeit im Kontext der Energiekontrollsysteme untersucht.

doi: 10.2313/NET-2015-03-1_08

5.1 Modifikationen des Verlaufsgraphen

Im folgenden Abschnitt wird der Graph, welcher den schützenswerten Verlauf darstellt, auf unterschiedliche Arten modifiziert. Um die nach Kapitel 4 sensiblen Informationen wie Minimum, Maximum, Peaks und deren Zeitpunkte zu schützen, existiert die Möglichkeit die Skalierung des Graphen bewusst zu manipulieren oder zu unterdrücken. Auf diese Weise sind keine Zeitpunkte von der x-Achse ablesbar, und die fehlende beziehungsweise verfälschte y-Achse verschleiert konkrete Werte.

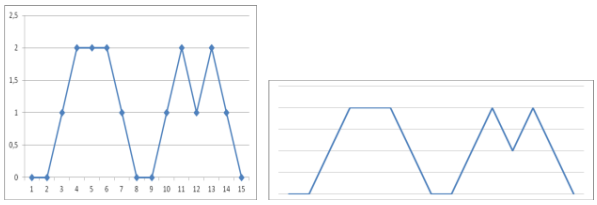


Figure 3. Graph mit Skalierung und ohne (rechts)

Obwohl sensible Daten geschützt wurden, enthält ein so modifizierter Verlauf (Figure 3) auch keine bei Auswertungen relevanten Informationen. Das Ändern der Skalierung ist reduzierbar auf den Pseudonymisierungsansatz bei Tabellen. Konkrete Werte werden durch mutmaßlich unsinnige oder falsche ersetzt und das Resultat schützt zwar die Privatsphäre der Nutzer, aber weist im Gegenzug keine wissenschaftliche Verwertbarkeit und Nutzbarkeit auf. Gerade bei Anwendungsfällen, die exakte Werte benötigen, etwa zu Abrechnungszwecken, ist von dieser Technik abzuraten. Auch bleiben Entwicklungen erkennbar.

Eine weitere Möglichkeit Interpretationen eines Verlaufsgraphen zu erschweren besteht im Flippen dieses Graphen mit einer Wahrscheinlichkeit p. Als Flippen definieren wir, dass beginnend von einem Ausgangswert zum Zeitpunkt t_0 jeder Anstieg zu einem Abstieg geändert wird. Der Betrag der Änderung bleibt dabei erhalten. Eine Zahlenfolge (1, 1, 3, 4) wird demnach in die Folge (1, 1, -1, -2) transformiert. Um eine triviale Rekonstruktion des tatsächlichen Verlaufs zu erschweren wird jeder Graph nur mit der Wahrscheinlichkeit p geflippt, mit der Wahrscheinlichkeit 1-p bleibt der ursprüngliche erhalten. Dennoch ist durch einen Hintergrundwissens-Angriff leicht herauszufinden ob es sich um eine modifizierte Ansicht handelt oder nicht. Dazu ausreichend ist bereits die Kenntnis der Art des Sensors beziehungsweise des durch ihn überwachten Verbrauchers. Der Stromverbrauch der meisten Arbeitsgeräte wird zu Beginn der Arbeitszeit vermutlich nicht nach unten gehen; diese Erkenntnis allein reicht in den meisten Fällen bereits aus um einen Graphen als geflippt oder original zu identifizieren. Auch ist diese Technik nicht für unterschiedliche Anwendungsfälle parametrisierbar.

Eine weitere Möglichkeit genaue Informationen zu verheimlichen besteht im Hinzufügen von Rauschen. Die variable Stärke des Rauschens kann dabei als Parameter der Granularität fungieren und entsprechend angepasst werden. Trotz des Rauschens können allerdings weiterhin Trends und Entwicklungen nachvollzogen werden; ist dies durch einen zu hohen Rauschanteil nicht mehr gegeben so kann auch davon ausgegangen werden, dass die Granularität zu gering ist und keinerlei Informationswert mehr enthalten ist. Auch ist Rauschen deshalb keine gute Option, da Aggregationen für Abrechnungen wie im Anwendungsfall des Buchhalters exakt sein müssen. Bedingt einsetzbar ist diese Technik allerdings in Fällen, in welchen das Zerlegen von zusammengesetzten Kurven verhindert werden soll. Kennt man charakteristische Muster einzelner Verbraucher, so ist man in der Lage aggregierte Zeitreihen durch Differenzbildung zu deaggregieren. Rauschen verhindert allerdings das saubere Entdecken von Mustern.

5.2 Abbilden auf Tabellen

Während der vorherige Abschnitt versuchte, Energieverläufe gemäß der Anwendungsfälle und Privatheitskriterien zu modifizieren, verfolgt dieses Unterkapitel Ansätze, welche die Verläufe auf Tabellen abbilden. Zielsetzung hierbei ist, im Anschluss daran die Techniken k-Anonymität, l-Vielfalt und t-Nähe verwenden zu können.

5.2.1 Pointer

In 5.1 wurde der Versuch unternommen die sensiblen Daten eines Verlaufs wie Ausschläge und genaue Zeitpunkte zu verheimlichen. Zu klären bleibt zudem die Frage, wie die Zugehörigkeit eines Verlaufs zu einem Sensor oder zu einer Person geschützt werden kann. Trivialerweise bietet sich dafür eine Tabelle bestehend aus der Identität des Nutzers und weiteren Attributen an. Diese weiteren Attribute definieren Pointer auf mit dieser Person gekoppelte Energieströme, zum Beispiel in seinem Büro, gebuchte Meetingräume, benutzte Gerätschaften etc. Diese Pointer weisen auf Dateien, in denen der Verlauf gespeichert ist.

Table 4. ID-1 identifiziert eine Person, ABC-12 einen mit dieser Person gekoppelten Verlaufsgraphen

Person	Büro	Meetingräume	Geräte
ID-1	ABC-12	DEF-34	GHI-45

Um diese Tabelle mittels z.B. l-Vielfalt zu anonymisieren, sollten statt eines expliziten Identifiers wie hier in *Table 4* Quasiidentifier aus mehrere Attributen verwendet werden, da die auf Unterdrückung von Teilen des Quasiidentifiers basierenden Techniken bei einem einzigen, eindeutigen Identifier nicht anwendbar sind. Je nach innerbetrieblicher Struktur bieten sich hierfür Büronummern, Geburtsjahre und ähnliches an.

Auf diese Weise kann eine Privatheit betreffend der Verwaltung der Kopplung von Personen und zeitlichen Entwicklungen gemäß der in Kapitel 3 erwähnten Techniken gewährleistet werden. Beachtet werden muss dabei allerdings auch, dass Angriffe mit zusätzlichem Hintergrundwissen unter Kollegen mit täglichem Umgang eine besondere Gefahr darstellen. Zu klären bleibt nun weiterhin die Frage wie die sensiblen Informationen innerhalb dieses Verlaufs geschützt werden können.

5.2.2 Charakteristika als Tabellenattribute

Nachdem wir in 5.2.1 eine Methode entwickelt haben den Zugang zu personenbezogenen Zeitserien gemäß geeigneter Kriterien zu verwalten wird nun eine Methode vorgestellt, die einen Schutz der Privatheit innerhalb dieser Daten mithilfe von Tabellen gewährleisten soll. Hierzu bedienen wir uns Eigenschaften eines Verlaufs, die diese Zeitserie zum Einen charakterisieren und zum Anderen in Tabellen abbildbar sind. Als solche erachten wir den Durchschnittswert, Gesamtsumme, sowie die Anzahl der

Zeitschritte, in denen der Wert innerhalb eines bestimmten Intervalls liegt. Auf diese Weise soll realisiert werden, bestimmten Rollen die laut Anwendungsszenario erforderlichen Daten zukommen zu lassen ohne die Möglichkeit einer zeitlichen Überwachung oder Verhaltensmustererkennung zu bieten. Durch diesen Ansatz können außerdem Trends und Entwicklungen verheimlicht werden wie später gezeigt werden wird.

Zur Veranschaulichung der Vorgehensweise betrachten wir folgenden Verlauf (Figure 4) und zwei ihn charakterisierende Tabellen (Table 5) in unterschiedlicher Granularität.

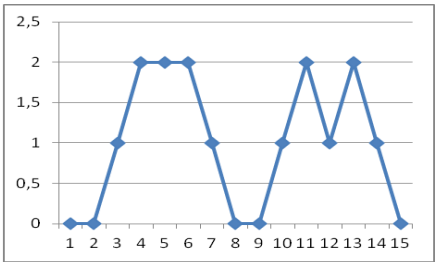


Figure 4. Beispieldaten eines Sensors

Table 5. Tabelle zu Graph aus *Figure 4*. Oben in höherer Auflösung, unten in niedrigerer

Durchschnitt	Summe	[0 ; 0,5[	[0,5 ; 1,5]	]1,5 ; 2,5]
1	15	5	5	5

Durchschnitt	Summe	<1	>= 1
1	15	5	10

Wie dieses Beispiel belegt, können durch Wahl unterschiedlicher Intervalle verschiedene Feinheitsgerade erreicht werden. Je nach Anwendungsfall ist die Darstellung also adaptierbar. Die sensiblen Daten Minimum und Maximum werden nicht offenbart; bei Wahl von geschlossenen Intervallen können diese lediglich einem Bereich zugeordnet werden, nicht aber exakt identifiziert werden. Bei Darstellung mit offenen Intervallen sind diese Eigenschaften sogar noch stärker verheimlicht. Auch werden schützenswerte Informationen über markante Ausschläge nicht preisgegeben. Gerade bei Profiling- und Überwachungsangriffen sind deren Zeitpunkte besonders interessant, doch diese Daten lassen sich aus obiger Darstellung weder ablesen noch berechnen. Außerdem wurden in Kapitel 4 erkennbare Trends als Gefahrenpotenzial identifiziert. Das folgende Beispiel belegt jedoch, dass auch Entwicklungen des Verlaufs mit dieser Technik nicht erkannt werden können. Sowohl *Figure 4* als auch die drei nachstehenden Graphen (Figure 5) werden auf die in *Table 5* zu findenden Tabellen abgebildet, wobei sie jedoch unterschiedlichste Trends darstellen.

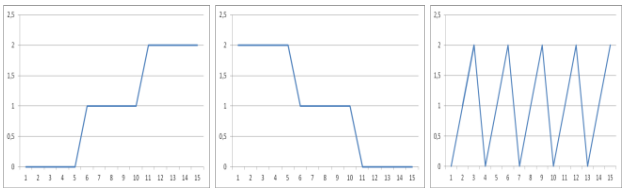


Figure 5. links ein monoton steigender Graph, mittig ein monoton fallender sowie rechts ein periodizierender Graph

Neben den nun erfüllten Anforderungen bezüglich der Verheimlichung von Entwicklungen sowie der Parametrisierbarkeit zur Bereitstellung unterschiedlicher Granularitäten ist es durch die Angabe von Durchschnittswerten auch für Public Displays geeignet. Ebenso können Buchhalter zu Abrechnungszwecken die angezeigte Gesamtsumme verwenden ohne Einblick in den tatsächlichen Verlauf der Zeitserie zu erhalten. Spezielle Methoden zur Berechnung dieser Summe werden in Abschnitt 5.3 vorgestellt. Doch nun kommen wir nochmals auf die in Kapitel 3 erläuterten Techniken zum Schutz sensibler Daten in Tabellen zurück und verbinden dieses Wissen mit der entwickelten Methode, Zeitserien auf Tabellen abzubilden. Hierzu wählen wir als Quasiidentifier den Verbund aus Raumnummer und Geburtsdatum, sensible Daten sind alle Charakteristika unserer Sensordaten analog zu obigem Beispiel. Eine so entstandene Beispieltabelle kann wie folgt aussehen (Table 6):

Table 6. Kombinierte Tabelle

Quasiidentifier		Sensible Daten				
Raum	*Geburt*	*Schnitt*	*Summe*	*<1*	*[1 ; 2]*	*>2*
05.5.23	1990	2,1	25	2	8	5
05.5.23	1964	1,9	23	6	4	5
05.5.12	1988	1,4	18	5	9	1

Auf diese Tabelle können nun die Techniken k-Anonymität, l-Vielfalt und t-Nähe fast ohne Adaption angewendet werden. Einem Buchhalter können 2-anonyme räumliche Aggregationen mit den Spalten Raum und Summe präsentiert werden, für Public Displays bieten sich l-vielfältige Durchschnittswerte an um einen Homogenitätsangriff abzuwehren. Die sehr gute Parametrisierbarkeit ergibt sich aus der Wahl der Intervalle und Intervallgrenzen, sowie der Entscheidung welche Spalten der sensiblen Daten veröffentlicht werden und der abschließenden Anwendung von k-Anonymisierung und artverwandten Techniken. Mit dieser Methode ist demnach eine Möglichkeit zur Privatsphäreerhaltung gefunden worden, die alle Privatheitskriterien wie oben erarbeitet erfüllt. Ein Nachteil allerdings liegt in der nötigen Diskretisierung kontinuierlicher Sensordaten um die Anzahl der Zeiteinheiten zu bestimmen, während jener der Messwert in einem bestimmten Intervall der Bildmenge lag.

5.3 Tatsächliche Aggregation

Dieser Abschnitt widmet sich dem aktuellen Forschungsthema, wie Summen von Privatsphäre-Restriktionen unterliegenden

doi: 10.2313/NET-2015-03-1_08

Summanden errechnet werden können. Interessant ist dies im Kontext der Energiekontrollsysteme besonders für den Anwendungsfall des Buchhalters. Neben diesem Aspekt wird zudem die Anpassbarkeit der Techniken relevant für ihre Anwendbarkeit sein. Hierzu untersuchen wir zunächst den in [9] verfolgten Ansatz. Dieser ist an das Verschlüsselungsverfahren nach Diffie-Hellman angelehnt.

In einem ersten Schritt wird ein zufälliger Generator g aus der zyklischen Gruppe G ausgewählt sowie n+1 geheime Zahlen s_i, sodass die Summe aller s_i null ist. Beginnend bei i=1 wird jedes s_i dem Teilnehmer i zugewiesen, welcher es als Secret Key verwendet. Der zu schützende Wert wird als x definiert, beziehungsweise x‘, falls ein Rauschen hinzugefügt wurde. Im Anschluss daran berechnet jeder Teilnehmer zum Zeitschritt t den Ciphertext c nach der Formel

$$c \leftarrow g^{x'} \cdot H(t)^{s_i}$$

wobei H(t) eine Hashfunktion auf G darstellt. Die verschlüsselten Werte werden nun dem Server übertragen, dieser berechnet

$$V \leftarrow H(t)^{s_0} \cdot \prod_{i=1}^{n} c_i$$

Mit $\prod_{i=1}^{n} (H(t)^{s_i}) = 1$ ergibt sich die Äquivalenz

$$V = g^{\sum_{i=1}^{n} x_i}$$

Nun kann der Server, dem g bekannt ist, die Summe aller ursprünglichen Werte mithilfe des diskreten Logarithmus berechnen. Die Erhaltung der Privatsphäre in diesem Verfahren wird analog zum Diffie-Hellman Algorithmus durch die Wahl einer geeigneten Gruppe G basierend auf einer genügend großen Primzahl p begründet. Mithilfe weniger Modifikationen kann diese Methode so adaptiert werden, dass statt Summen auch Produkte sensibler Informationen aggregiert werden können.[9] Durch die Verschlüsselung können weder der Server noch ein potenzieller Man in the Middle die sensiblen Summanden auslesen. Nach Durchlauf des Protokolls steht dem Server lediglich die aggregierte Summe zur Verfügung. Obwohl das Verfahren also Angriffe gegen den Datenschutz zu unterbinden vermag, wird dennoch zum Austausch der Schlüssel eine vertrauenswürdige Verbindung benötigt.

Ein weiterer Ansatz mit ähnlichem Ziel wird in [8] verfolgt. Auch hier wird mit einer Initiationsphase und der Wahl eines Generators g begonnen. Zunächst wird ein privater Schlüssel λ festgelegt, und jedem User u wird ein λ_u zugewiesen, sodass λ die Summe aller λ_u darstellt. Anschließend addiert jeder Nutzer zu seinem geschützten Wert x_u eine nur ihm bekannte Zufallszahl r_u und verschlüsselt das Resultat, da Rauschen alleine keinen genügenden Schutz darstellt. Verwendet wird dazu das Paillier-Kryptosystem. Der zentrale Server berechnet nun das Produkt aller übertragenen Werte, welches in der verschlüsselten Darstellung der Summe aller (x_u + r_u) besteht. Nach dem Distributivgesetz gilt $c = \sum_{u=1}^{|User|}(x_u + r_u) = \sum_{u=1}^{|User|} x_u + \sum_{u=1}^{|User|} r_u$.

Nachdem der Server aber die r_u nicht kennt, sendet er c an alle Teilnehmer zurück. Diese berechnen nun Antworten basierend auf ihrem persönlichen Schlüssel und ihrer Zufallszahl r und schicken diese Antwort c’$_u$ an den Server zurück. Dabei gilt $c'_u = c^{\lambda_u} \cdot g^{-r_u \cdot \lambda}$. Das Produkt dieser c’$_u$ berechnet der Server um anschließend die endgültige Entschlüsselung der gesuchten Summe wie folgt als $\frac{L(c' \bmod m^2)}{L(g^{\lambda} \bmod m^2)}$ zu berechnen. Dabei bezeichnet m eine Primzahl, so dass alle verschlüsselten Nachrichten aus {0,1,…,m-1} stammen. Auch wird definiert $L(z) = \frac{z-1}{m}$.

Wir fassen diese Schritte wie folgt zusammen: Um die Summe aller x_u zu berechnen fügt jeder Nutzer seinem Wert ein Rauschen r_u hinzu. Da das Rauschen alleine keinen genügenden Schutz gewährleistet, verschlüsselt jeder Nutzer die Summe aus seinem Wert und Rauschen. Aufgrund der Verschlüsselung kann der Server lediglich die Verschlüsselung der Summe aller übertragenen (x_u + r_u) berechnen. Dieser verschlüsselte Wert wird anschließend zu den Klienten zurückgesandt, welche diesen dazu benutzen basierend auf ihrem Secret Key Informationen zur Entschlüsselung der Gesamtsumme bereitstellen.

Für einen formalen Beweis dieser Methode sei auf [8] und [10] verwiesen. Der Vorteil dieser Variante liegt in der verteilten Entschlüsselung der Daten, so kann ein Angreifer selbst dann keine Informationen gewinnen, wenn es ihm gelingen sollte einen der Teilnehmer zu infiltrieren. Andererseits liegt in der dezentralen Abwicklung auch eine große Schwäche der Variante, da so ein falscher, fehlender oder zu langsam übertragener Wert zu einem Misserfolg führen kann. In der Konsequenz wird also ein permanenter Online-Status aller Nutzer erwartet.

6. Bewertung hinsichtlich Anwendungsfälle

Um eine Empfehlung hinsichtlich einer Entscheidung für eine der erläuterten Datenschutztechniken abgeben zu können, befasst sich dieser Abschnitt nochmals mit den Vor- und Nachteilen jener und evaluiert sie bezüglich der gegebenen Anwendungsfälle. Diese bestanden zum Einen in der Anforderung, dass Nutzer die von ihnen verursachten Energiedaten in höchster Auflösung einsehen können, die anderer Teilnehmer allerdings nur mit deren Einverständnis. Die Daten in feinster Auflösung müssen folglich unverfälscht zur Verfügung stehen. Zum Anderen sollen Public Displays in mittlerer Granularität zudem die Energietransparenz eines Gebäudes erhöhen, auch sind räumliche Aggregationen zu Abrechnungszwecken sinnvoll. Da folglich verschiedene Feinheitsgrade der Daten zur Verfügung gestellt werden müssen, ist außerdem die Parametrisierbarkeit einer Technik ein wichtiger Punkt. Eine solche Anpassbarkeit stellen zwar k-Anonymität, l-Vielfalt und t-Nähe bereit, haben allerdings den Nachteil des sinkenden Informationswerts der Daten bei steigendem Datenschutz. Außerdem basieren diese Methoden auf der Tatsache, dass Daten in Tabellenform vorliegen, was hier nicht der Fall ist. Nachdem die Daten des Energiekontrollsystems allerdings durch die in 5.2.2 beschriebene Methode in diese Darstellungsweise überführt wurden, sind jene Verfahren

durchaus anwendbar. Der Einsatz dieser Kombination wird für die meisten Anwendungsfälle empfohlen. Für das Szenario eines Buchhalters, der lediglich summierte Verbräuche in bestimmten Gebäudeteilen benötigt, legen wir dem Anwender nahe, eines der Verfahren aus 5.3 zu verwenden. Diese garantieren eine Summierung der Werte ohne Einblick in die Einzelverbräuche zu gewähren. Der Einblick eines Nutzers in seine eigenen Energiedaten und der Zugang des Energiemanagers zu diesen können in einer Tabellenform angelehnt an 5.2.1 verwaltet werden, bei der Links auf feingranulare Daten zur Verfügung gestellt werden.

7. ZUSAMMENFASSUNG

Mit dem derzeit steigendem Energiebewusstsein vieler Menschen erhielten auch Energiekontrollsysteme Einzug in Privathäuser ebenso wie in gewerbliche oder behördliche Gebäude. Diese Systeme sammeln kontinuierlich Energieverbrauchsdaten durch eine Vielzahl an Sensoren. Dabei soll es jedem Nutzer möglich sein, von ihm verursachte Energiedaten in höchster Auflösung einsehen zu können, die anderer Teilnehmer allerdings nur mit deren Einverständnis. Bestimmte Rollen wie zum Beispiel ein Buchhalter sollen zudem aggregierte Informationen zu Abrechnungszwecken erhalten können, ebenso sollen Public Displays bereitgestellt werden können. Es soll außerdem nicht möglich sein, mithilfe der freigegebenen Daten Teilnehmer des Energiekontrollsystems zu überwachen oder ein Verhaltensmuster jener zu erkennen, weshalb markante Stellen im Verlaufsgraphen und deren zeitlicher Zusammenhang nicht offenliegen dürfen. Etablierte Methoden zum Schutz sensibler Daten, beispielsweise aus dem Bereich der medizinischen Statistik, stützen sich auf eine Abbildung der Informationen in Tabellen. Hierbei bezeichnet man einen Verbund nicht sensibler Daten, die ein Individuum identifizieren können, als Quasiidentifier. Ein verbreiteter Ansatz ist nun eine Unentscheidbarkeit bei der Zuordnung eines solchen Quasiidentifiers zu einem geschützten Datum zu erreichen. Eine solche Methode namens k-Anonymität stellt eine Tabelle bereit, bei der jeder Quasiidentifier von mindestens k-1 anderen Zeilen nicht zu unterscheiden ist. Dennoch ist in manchen Fällen eine Zuordnung einer Person zu einem sensiblen Eintrag möglich, falls entweder zusätzliches Hintergrundwissen vorliegt oder die sensiblen Daten innerhalb der k nicht unterscheidbaren Einträgen identisch sind. Um dies zu verhindern wird die Technik l-Vielfalt dargestellt, die zusätzlich eine Heterogenität innerhalb der durch k-Anonymisierung entstandenen Äquivalenzblöcke gewährleistet. Allerdings sagt l-Vielfalt lediglich etwas über die Unterschiedlichkeit der sensiblen Daten aus, nicht jedoch über deren Ähnlichkeit. So ist ein Angriff auf die Privatsphäre dennoch möglich, falls zwar eine große Vielfalt an Werten herrscht, aber diese alle einer gewissen Überkategorie zuzurechnen sind. Somit kann leicht auf diese Überkategorie geschlossen werden, welche allerdings ebenfalls als schützenswert erachtet wird. Um dies zu verhindern fordert die Technik t–Nähe die Unterschreitung einer gewissen Distanz zwischen der Verteilung der sensiblen Werte innerhalb nicht unterscheidbarer Quasiidentifier und ihrer Verteilung über die Gesamtheit aller Einträge.

Obwohl durch diese Schritte der Schutz der Privatsphäre zwar stark zunimmt, schwindet jedoch durch verstärkte Restriktionen und vermehrte Unterdrückung von Teilen des Quasiidentifiers der Nutzen und wissenschaftlicher Informationswert des Outputs. Auch sind diese Techniken lediglich auf Tabellen anwendbar. In einem Energiekontrollsystem sind hingegen vor allem Energieverläufe interessant, weshalb im Anschluss zuerst eine Technik vorgestellt wird solche Verläufe in Tabellen zu verlinken, bevor eine Methode entwickelt wird diese Verläufe auf Tabellen anhand ihrer Charakteristika abzubilden. Sinnvoll ist dies deshalb, da hierdurch wiederum die etablierten Methoden k-Anonymität und artverwandte Techniken angewandt werden können. Um Privatsphäre-erhaltend Summen zu generieren werden zwei Methoden vorgestellt, welche dies ohne Kenntnis der tatsächlichen Summanden ermöglichen. Ein Einsatz dieser Verfahren wird im Anwendungsfall eines Buchhalter als sinnvoll erachtet, da ihm somit lediglich Aggregationen zu Abrechnungszwecken zur Verfügung stehen und keine Einzeldaten bekannt werden. Für den Anwendungsfall des Energiemanagers ebenso wie für die Einsicht der Nutzer in die eigenen Daten wird ein Zugang zu einer Tabelle gemäß der in 5.2.1 dargestellten Technik empfohlen, bei der Links auf feingranulare Daten zur Verfügung gestellt werden. Für weitere Anwendungsszenarien wird die Verwendung einer Charakteristika-basierten Tabellendarstellung der Verläufe analog zu 5.2.2 nahegelegt, welche anschließend mithilfe der Verfahren zum Schutz von Tabellendaten aggregiert werden kann. Mit dem Spiel EQ wird zudem ein spielerischer Wettkampf präsentiert, welcher den Nutzern eine Einordung des eigenen Energieverbrauchs bereitstellt ohne sensible Daten anderer offenzulegen.

8. REFERENZEN

[1] U.S. Department of Health and Human Services, “De-identifying protected health information under the privacy rule”, 2007. *URL: http://privacyruleandresearch.nih.gov/pr_08.asp#8a, zul. aufgerufen am 28.08.2014*

[2] Sweeney, Latanya. "k-anonymity: A model for protecting privacy." *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 10.05 (2002): 557-570.

[3] Li, Ninghui, Tiancheng Li, and Suresh Venkatasubramanian. "t-Closeness: Privacy Beyond k-Anonymity and l-Diversity." *ICDE*. Vol. 7. 2007.

[4] Klaus Kuhn, „Datenschutz in der Medizin“, *Vorlesungsunterlagen zur Veranstaltung Medizin II (Krankheitslehre, klinische Propädeutik, Einführung in die Medizinische Informatik)*, IMSE-TUM, 2014 URL:https://www.moodle.tum.de/pluginfile.php/389365/mod_resource/content/1/med2_ss14_v10a.pdf

[5] Machanavajjhala, Ashwin, et al. "l-diversity: Privacy beyond k-anonymity." *ACM Transactions on Knowledge Discovery from Data (TKDD)* 1.1 (2007): 3.

[6] Bundesdatenschutzgesetz, Bundesministerium der Justiz und für den Verbraucherschutz. *URL: http://www.gesetze-im-internet.de/bdsg_1990/BJNR029550990.html*

[7] Zhu, Ye, Yongjian Fu, and Huirong Fu. "On privacy in time series data mining."*Advances in Knowledge Discovery and Data Mining*. Springer Berlin Heidelberg, 2008. 479-493.

[8] Rastogi, Vibhor, and Suman Nath. "Differentially private aggregation of distributed time-series with transformation and encryption." *Proceedings of the 2010 ACM SIGMOD*

International Conference on Management of data. ACM, 2010.

[9] Shi, Elaine, et al. "Privacy-Preserving Aggregation of Time-Series Data." *NDSS*. Vol. 2. No. 3. 2011.

[10] Rastogi, Vibor., and Nath, Suman. “Differentially private aggregation of distributed time-series with transformation and encryption”. *Tech.Rep. MSR-TR-2009-186*, Microsoft Research, 2009. Extended version of [8]

Internet Science – Critical Infrastructures

Caterina Wanka
Betreuer: Dr. Heiko Niedermayer
Seminar Future Internet WS2014/15
Lehrstuhl Netzarchitekturen und Netzdienste
Fakultät für Informatik, Technische Universität München
Email: caterina.wanka@tum.de

KURZFASSUNG

Kritische Infrastrukturen versorgen uns mit dem Wasser, das wir trinken, dem Strom, welchen wir im Alltag benötigen, den Transportmitteln, welche uns zur Arbeit bringen, und den Kommunikationssystemen, über welche wir mit Freunden und Familie in Kontakt bleiben. Insbesondere Energieversorgungssysteme sind in der heutigen Gesellschaft von großer Bedeutung. Eine Störung oder ein Ausfall dieses Systems kann katastrophale Auswirkungen mit sich ziehen. Insbesondere Stromnetze verdeutlichen die Komplexität und Vielfältigkeit kritischer Infrastrukturen. Auf Grund ihrer Bedeutung in unserer heutigen Gesellschaft sind sie oft Zielscheibe von Angriffen unterschiedlichster Art. Jedoch gibt es eine Vielzahl von Schutzmaßnahmen, welche schädigenden Ereignissen entgegenwirken können.

Schlüsselworte

Kritische Infrastruktur – Energieversorgungssystem –Informations- und Kommunikationstechnik (IKT) – Critical Infrastructure Protection (CIP) – Cyber-Sicherheit

1. EINLEITUNG

"[Critical infrastructures] are the foundations of our prosperity, enablers of our defense, and the vanguard of our future. They empower every element of our society. There is no more urgent priority than assuring the security, continuity, and availability of our critical infrastructures." [1]

Moderne Industrienationen sind auf komplexe Infrastrukturen angewiesen. Wirtschaft und Gesellschaft funktionieren nur, wenn die grundlegende Versorgung gesichert ist. Ein Ausfall oder eine Manipulation und Beeinträchtigung dieser Systeme über einen längeren Zeitraum oder auf einer größeren Fläche würde weitreichende Folgen nach sich ziehen. Deswegen stehen kritische Infrastrukturen und im besonderen Maße deren Schutz weltweit im Mittelpunkt nationaler Regierungsaktivitäten.

Die Energieversorgung ist ein zentraler Bereich kritischer Infrastrukturen, da das Energieversorgungssystem neben seiner weiten geographischen Verbreitung, zudem der Schlüssel zu den meisten sozialen Aktivitäten ist. Ausfälle oder Störungen des Energieversorgungssystems würden sich extrem und unmittelbar auf die anderen Sektoren und somit auf Staat, Wirtschaft und Gesellschaft auswirken.

Diese Arbeit schafft einen Einblick in die Thematik ‚Kritische Infrastrukturen'. Am Beispiel des Sektors der Energieversorgung werden die Akteure und deren Zusammenspiel im System kritischer Infrastrukturen näher untersucht. Daran anknüpfend wird ein Überblick über die ‚Critical Infrastructure Protection' verschafft, indem Ursachen für das Scheitern von kritischen Infrastrukturen untersucht und abschließend Schutzmaßnahmen vorgestellt werden.

2. WAS SIND KRITISCHE INFRASTRUKTUREN?

Der Begriff ‚Kritische Infrastrukturen' ist in der Wissenschaft, wie auch in der Politik vieldiskutiert. Auf Grund des weitreichenden Begriffsumfangs wird im Folgenden zusätzlich zu einer reinen Begriffsbestimmung auch eine Einteilung der zu untersuchenden Systeme in Sektoren vorgenommen.

2.1 Definition

Unterschiedliche Ansätze und Interpretationen führen zu unterschiedlich weitgreifenden Definitionen, weswegen eine unabhängige Untersuchung der beiden Worte ‚kritisch' und ‚Infrastruktur' sinnvoll erscheint, um schließlich zu einer umfassenden Definition zu gelangen.

Infrastruktur: Eine Infrastruktur ist die Gesamtheit an Elementen, welche zur Ausführung einer bestimmten Dienstleistung notwendig ist. Folglich ist es von der Perspektive abhängig, welches System für ein anderes eine Infrastruktur darstellt. Infrastrukturen können auch innerhalb von anderen Infrastrukturen existieren. Dieser Ansatz lässt sich gut anhand der beiden Systeme Energieversorgung und Informations- und Kommunikationstechnik erklären. Vereinfacht gesagt benötigt Stromübertragung Netzwerke und Netzwerke benötigen andererseits Strom. Aus Sicht des jeweiligen Systems stellt das andere wiederum Teil seiner Infrastruktur dar. Für diese Arbeit wird auf Grund der gesellschaftspolitischen Relevanz des Themas von der abstrakten Perspektive eines Staates ausgegangen. Obwohl jede Regierung den Rahmen zu betrachtender Objekte unterschiedlich definiert, erkannte die Organisation für wirtschaftliche Zusammenarbeit und Entwicklung (OECD), dass die meisten Staaten den Begriff ‚Infrastruktur' aus einer eher weiten Perspektive betrachten[2]. Aus dieser Sicht umfasst der Begriff physische Objekte, wie zum Beispiel Telefonleitungen, Stromnetze und Gasleitungen. Hinzu kommen kritische Informationsinfrastrukturen. Diese umfassen wiederum physische Systeme, unter anderem bestehend aus Highspeed- oder Breitbandnetzwerken. Der andere Teil umfasst die immaterielle Komponente, in Form von Daten und Software, welche eingebettet in Computersystemen, physische Infrastrukturen bedienen.

doi: 10.2313/NET-2015-03-1_09

Kritisch: ‚Kritisch' sind Infrastrukturen für einen Menschen, sobald sie ernstzunehmend für die Erhaltung seiner Lebensqualität notwendig sind. Das heißt, gegensätzlich betrachtet, bei dessen Ausfall das Leben eines Menschen gefährdet wäre oder geschädigt werden würde. Aus Sicht eines Staates sind folglich Infrastrukturen ‚kritisch', sofern sie einen essentiellen Beitrag zum wirtschaftlichen und sozialen Gemeinwesen des Landes leisten. Das bedeutet, sobald bei deren Ausfall oder Beeinträchtigung nachhaltig und/ oder weitreichend wirkende Versorgungsengpässe, erhebliche Störungen der öffentlichen Sicherheit oder andere dramatische Schäden eintreten würden[3], wird eine Infrastruktur auf nationaler Ebene als ‚kritisch' eingestuft.

Auf Grund der sozialpolitischen Relevanz dieses Begriffs haben viele Staaten eigene Definitionen beziehungsweise Beurteilungskriterien formuliert, die im Rahmen ihrer nationalen Schutzprogramme für kritische Infrastrukturen verwendet werden. Die Vereinigten Staaten zum Beispiel bezeichnen kritische Infrastrukturen als Systeme oder Güter, physisch oder virtuell, welche für die Vereinigten Staaten dermaßen entscheidend sind, dass eine Störung beziehungsweise Zerstörung dieser Systeme und Güter eine lähmende Auswirkung auf die nationale Sicherheit mit sich ziehen würde[4]. Damit legen sie den Fokus auf den Schutz der nationalen Sicherheit und bewerten den Einfluss einer Infrastruktur auf das Gemeinwesen nach dem Maß der Auswirkungen auf ihre nationale Sicherheit. Empfehlenswert ist es zudem den Ansatz von der entgegengesetzten Seite zu betrachten und ein Gut oder einen Service als kritisch zu betrachten, sobald sie für die Aufrechterhaltung von lebenswichtigen sozialen Funktionen essentiell sind.

Somit wird der Grad an Bedeutung einer Infrastruktur nicht an den hypothetischen Auswirkungen im Falle eines Ausfalls, sondern an dem Beitrag zum nationalen Gemeinwesen gemessen. Beide Herangehensweisen führen im Durchschnitt zu den gleichen Ergebnissen, jedoch lassen sich durch die letztere mehr unterstützende Funktionen kritischer Infrastrukturen in die Betrachtungsweise mit aufnehmen.

2.2 Sektoreinteilung

Auch eine Einteilung kritischer Infrastrukturen in Sektoren ist Aufgabe der Regierung und folglich sind auch hier eine Vielzahl an unterschiedlichen Ansätzen vorzufinden. Kanada unterteilt kritische Infrastrukturen zum Beispiel in zehn Branchen, die Vereinigten Staaten dagegen haben achtzehn Sektoren. Alles in allem sind die Einteilungen trotz unterschiedlicher Anzahl an Sektoren vergleichbar. Als Ansatz für diese Arbeit dient die Einteilung des deutschen Bundesamts für Bevölkerungsschutz und Katastrophenhilfe[5] in neun Sektoren:

- Energie
- Gesundheit
- Staat und Verwaltung
- Ernährung
- Transport und Verkehr
- Wasser
- Finanz- und Versicherungswesen
- Informationstechnik und Telekommunikation
- Medien und Kultur

Hervorzuheben ist insbesondere der Bereich der Informationstechnik und Telekommunikation (IKT). Wie bereits im Rahmen der Definition erläutert, nennt man diesen Teilbereich der jeweiligen Sektoren kritischer Infrastrukturen in der Wissenschaft auch ‚kritische Informationsinfrastruktur'[6]. Auf eine solche Betrachtung wird in den meisten Ländern verzichtet. Ein beliebter Ansatz von Staaten ist die Zuordnung des Gefüges kritischer Informationsinfrastrukturen in den Sektor der IKT, welcher in engen Interdependenzen zu den restlichen Sektoren steht. Die zunehmende Durchdringung aller Lebens- und Arbeitsbereiche durch IKT bestimmt jedoch maßgeblich unseren technologischen Fortschritt und lädt zu einer Betrachtung der Informationsinfrastruktur als Teilbereich der einzelnen Sektoren ein. Ein Paradebeispiel hierfür ist das Energieversorgungssystem. Informationsinfrastrukturen stellen hierbei die Schnittstelle dar, welche im Rahmen der Kooperation und Koordination der miteinander verbundenen Stromnetze die Kommunikation und den Datenaustausch untereinander ermöglichen.

3. STRUKTUR UND FUNKTIONSWEISE VON KRITISCHEN INFRASTRUKTUREN

Die wachsende Vernetzung von Dienstleistungen, Infrastrukturen und Prozessen hat zur Folge[7], dass weite Bereiche des gesellschaftlichen und wirtschaftlichen Lebens von funktionierenden, robusten Infrastrukturen abhängig sind.

Stromnetzsysteme bestehen heutzutage aus einer Vielzahl an aufeinander wirkenden nationalen Systemen. Diese sind vorwiegend über weite geographische Gebiete miteinander verbunden. Am Beispiel des europäischen Netzes betrifft dies sogar den gesamten Kontinent. Kontroll- und Kommunikationszentralen dienen zum Austausch von Daten und Anweisungen zwischen den regional und national abgegrenzten Netzsystemen. Die Elemente innerhalb der Stromnetzsysteme lassen sich folglich, wie in Abbildung 1 verdeutlicht, laut den Wissenschaftlern Negenborn, Lukszo und Hellendoorn[8] in drei Ebenen einordnen:

- **Physische Ebene**: Energieerzeugung und -übertragung
- **Entscheidungsebene**: Organisatorische und menschliche Entscheidungen
- **Cyber-Ebene**: Übertragung von Informationen und Befehlen

3.1 Physische Ebene

In Stromnetzen besteht die physische Ebene aus der Netzwerkhardware. Das Joint Research Centre des Institutes für Energie und Transport[9] beschreibt diese als Anordnung technischer Bauelemente, welche interagierend den Prozess von der Stromerzeugung bis hin zur Lieferung des Stroms an den Endverbraucher realisieren.

Der Energiefluss ist unidirektional: von den zentralisierten Erzeugern bzw. Kraftwerken hin zu den Verbrauchern, welche sich auch nach ihrer Menge an verbrauchtem Strom unterscheiden lassen. Dieser Prozess kann in vier Subsysteme unterteilt werden. Abbildung 1 veranschaulicht im oberen Bereich (‚Physical Layer') eine vereinfachte Struktur eines Elektrizitätsnetzwerkes bestehend aus vier Spannungsebenen[10]:

doi: 10.2313/NET-2015-03-1_09

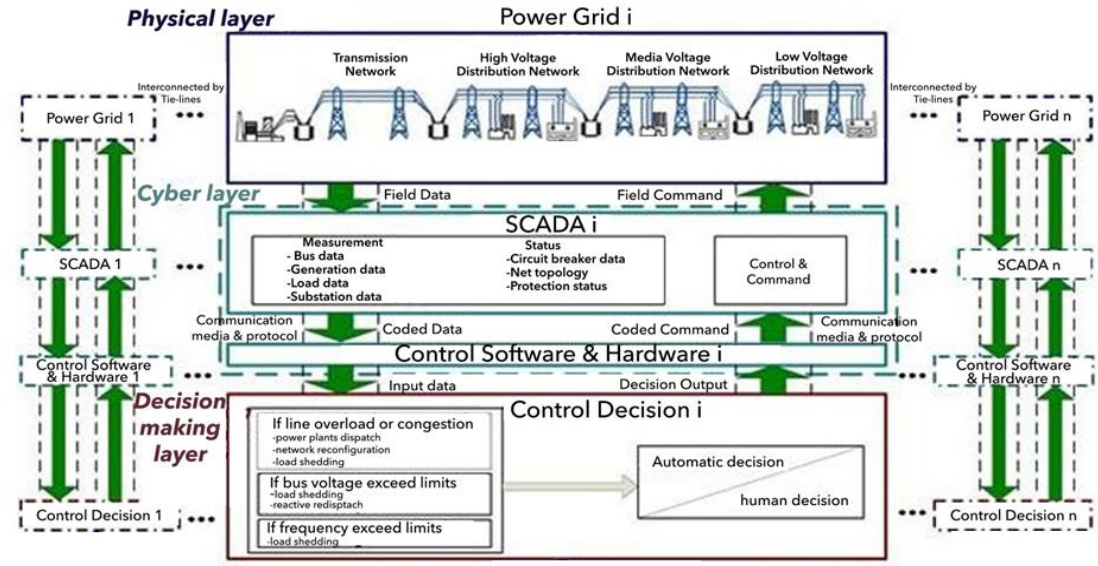


Abbildung 1. Die drei Ebenen des Energieversorgungssystems [8]

Dem Übertragungsnetz (Höchstspannungsnetz HÖS), dem Hochspannungsnetz (HS), dem Mittelspannungsnetz (MS) und dem Niederspannungsnetz (NS).

Die Übergänge zwischen den verschiedenen Spannungsebenen werden durch Transformatoren (Aufwärts- beziehungsweise Abwärtstransformatoren) in Umspannwerken oder Ortsnetzstationen realisiert.

Die Stromerzeugung findet heutzutage überwiegend in großen Kraftwerken, wie zum Beispiel Kohlekraftwerken oder Atomkraftwerken, statt. Ressourcen hierfür sind[11] vorwiegend Kohle und Kernenergie. Ein konstant wachsender Anteil der Stromerzeugung besteht zudem aus erneuerbarer Energie, welche in Windkraftanlagen, Solaranlagen etc. erzeugt wird. Der Betrieb in den Kraftanlagen wird zentral überwacht und die Einspeisung der erzeugten Energie in das Stromnetz koordiniert.

Über Aufwärtstransformatoren wird die erzeugte Energie in das Übertragungsnetz (‚Transmission Network') verbreitet. Auf der Höchstspannungsebene können lediglich Großkraftwerke mit Leistungen bis zu 700 MW einspeisen, kleinere Kraftwerke sowie Windkraftanlagen speisen auf Hoch- und Mittelspannungsebene ein.

Über die Höchstspannungsebene ist das deutsche Stromnetz einerseits in das europäische Verbundnetz UCTE (Union for the Coordination of Transmission of Electricity)/ ENTSO-E (European Network of Transmission System Operators for Electricity) für Mittel- und Südeuropa eingebunden. Darüber hinaus hat es über Hochspannungsleitungen Verbindungen in das skandinavische Verbundnetz NORDEL[10]. Zudem gibt es einige industrielle Verbraucher, die auf Grund der großen Menge an benötigtem Strom direkt über das Übertragungsnetzwerk versorgt werden.

Abwärtstransformatoren wandeln den Strom auf eine niedrigere Spannungsebene um, um ihn in das Verteilnetzwerk (‚Distribution Network') zu übertragen.

Die Vielzahl an Stromleitungen werden durch Knoten, sogenannte ‚Busse', miteinander verbunden. Im HS werden diese Busse durch Umspannwerke verkörpert. Umspannwerke wandeln folglich nicht nur den übertragenen Strom in eine niedrigere Spannungsebene um, sondern sind generell für die Kontrolle und Regulierung der Stromflüsse zwischen den Leitungen verantwortlich. Die Versorgungssysteme werden übergeordnet von einer nationalen Aufsichtsbehörde reguliert. Auf europäischer Ebene wurden die unterschiedlichen nationalen Vorschriften durch die ENTSO-E angepasst. Bezüglich des Verantwortungsbereichs identifizierten die Wissenschaftler Bompard et al.[12] Konsequenzen vor allem im Bereich der Busse. Folglich lassen sich diese ‚Übertragungsknoten' im Hinblick auf ihre Zugehörigkeit und ihres physischen Verhaltens in vier Kategorien unterteilen:

Busse des Übertragungsnetzwerkes werden durch sogenannte Transmission Stations (TS) verkörpert, welche direkt dem Transmission System Operator (TSO), dem sogenannten Übertragungsnetzbetreiber, angehören und durch diesen betrieben werden.

Power Plants (PP) sind Kraftwerke, welche verschiedenen, untereinander stark konkurrierenden Unternehmen angehören. Ein Unternehmen kann mehrere Kraftwerke besitzen, welche nicht an denselben Bussen des Netzwerkes angebunden sein muss.

Ein weiterer Bus-Typ sind die Distribution System Feeders (DS). Durch diese Busse können Betreiber über ein abgegrenztes Verteilnetzwerk als Monopolist verfügen.

Von Large Users (LU)-Bussen werden Verbraucher direkt versorgt, sobald ihre Nachfrage an Strom über 5 MW liegt.

Der Mensch ist als Akteur an sich auf physischer Ebene nicht direkt für den Betrieb eines modernen Stromnetzes notwendig. Neben automatisierten Vorrichtungen ist er in diesem Bereich zum Beispiel für die Ausführung von Wartungsarbeiten an den Elementen des Energieversorgungssystems zuständig.

doi: 10.2313/NET-2015-03-1_09

3.2 Entscheidungsebene

Der Betrieb von Kraftwerken wird auf der höchsten Ebene durch einen Layer der Entscheidungsfindung bewerkstelligt. Die Entscheidungsebene umfasst die organisatorischen und menschlichen Entscheidungen[8], welche den Betriebsablauf des Stromnetzsystems betreffen. Charakterisierend hierfür ist das Zusammenspiel automatischer Kontrollinstanzen und menschlicher Entscheidungen, wie auch in Abbildung 1 dargestellt. Die höchste Ebene im Energieversorgungssystem besitzt somit eine lenkende Funktion. Die Struktur dieser Ebene basiert auf einer klaren Trennung der Bereiche Steuerung und Monitoring. Abbildung 2 verdeutlicht die hierarchische Anordnung der Akteure der Entscheidungsebene, welche durch Kontrollzentren verkörpert werden.

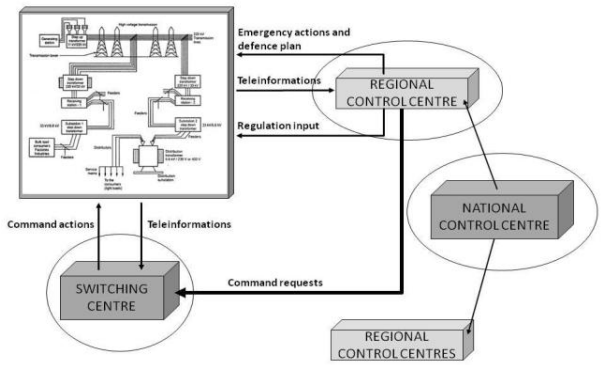


Abbildung 2. Strukturüberblick der Kontrollzentren [12]

Die Steuerung wird über die sogenannten Switching Centers (SC) ausgeführt, welche für die Übermittlung von Anweisungen an das Stromnetz zuständig sind. Diese Befehle resultieren letztendlich in Konfigurationsänderungen der Bauelemente (Leistungsschalter, Sicherungslasttrenner, Stufenschaltwerk etc.) des Stromnetzes.

Das Monitoring ist Aufgabe der Regional Control Centers (RCC). Sie kontrollieren den Netzwerkstatus, verwalten dessen Inputdaten und übermitteln an die Betreiber der SCs Steuerungsanfragen für das Stromnetz.

Über diesen beiden Institutionen steht das National Control Center (NCC), welches die Handlungen der RCCs überwacht und nach vorgegebenen Regelungen die Stromflüsse zu anderen Netzen koordiniert.

Kontrollzentren können über zwei Arten von erweiterter Anwendungssoftware verfügen. Beide übernehmen jeweils voneinander getrennte Aufgaben der Verwaltung und Steuerung eines Energieversorgungssystems. Systeme, die mit einer Software für analytische Funktionen im Bereich des Netzwerkbetriebs ausgestattet sind (u.a. Zustandsschätzungen, Netzwerkanalysen, Erzeugungssteuerung), nennt man Energy Management Systems (EMS). Das sogenannte Business Management System (BMS) ist der Teil der Kontrollzentren, welcher für die kommerziellen Anwendungen zuständig ist[13]. Durch Human-Machine-Interfaces (HMI) werden innerhalb der jeweiligen Kontrollzentren Schnittstellen zwischen dem Betreiber (‚Human') und den EMS- und BMS-Systemen hergestellt.

Alles in allem ist in der Entscheidungsebene der Mensch Hauptakteur, verkörpert durch das Personal der Koordinations- und Kontrollzentren. Neben der stetig wachsenden Automatisierung von Anweisungen, ist er grundsätzlich für die Initialisierung von organisatorischen und Sicherheitsmaßnahmen verantwortlich. Seine Anordnungen werden durch HMIs über EMS- oder BMS-Systeme auf eine Datenebene übersetzt und an die physische Ebene übermittelt.

3.3 Cyber-Ebene

Wie in Abbildung 1 dargestellt, wird durch die Cyber-Ebene eine bidirektionale Kommunikation zwischen der physischen Ebene und der Entscheidungsebene gewährleistet. Die Datenübertragung erfolgt in Richtung der Entscheidungsebene und die Kontrollhandlungen hin zur physischen Ebene.

Sogenannte Remote Terminal Units (RTU) stellen, wie Abbildung 3 veranschaulicht, die Schnittstelle zwischen den Netzwerk-Bussen in der physischen Ebene zur Cyber-Ebene dar[8]. RTUs sind einfache Bauelemente, welche mit einem Mikroprozessor und einer bestimmten Menge an digitalen und analogen Input/ Output-Kanälen ausgestattet sind. Manche Busse sind direkt und ausschließlich zu einem ausgewählten Bus verbunden, während dagegen andere gruppiert werden, um alle Informationen von einer Vielzahl an Bussen am gleichen Ort zu konzentrieren und sie für eine RTU zugänglich zu machen. Ein Kraftwerk ist in der Lage die Informationen von anderen Kraftwerkanlagen zu verwalten. Aus diesem Grund sind PPs mit einem fest zugehörigen RTU ausgestattet, insbesondere sobald Anlagen mit einer großen Menge an Informationen versorgt werden müssen.

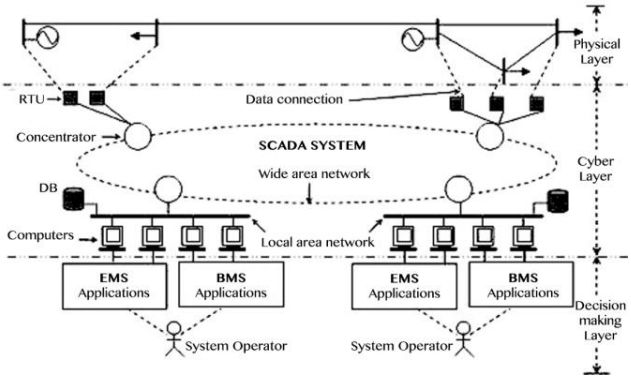


Abbildung 3. Cyber-Ebene - Kontrollsystem [12]

RTUs benötigen eine sichere bidirektionale Kommunikation zwischen den Einrichtungen der RCCs und SCs und der Netzwerkhardware. Im Rahmen dieser Kommunikation werden über das Supervisory Control and Data Acquisition (SCADA) System die Kontrollzentren mit einer Menge an Informationen über den Netzwerkstatus versorgt. Das SCADA System verkörpert somit eine ‚Cyber-Brücke'[8] zwischen dem physischem System und den Kontrollzentren, die über die EMS- und BMS-Anwendungen mit dem Kommunikationssystem verbunden sind.

4. CRITICAL INFRASTRUCTURE PROTECTION

Das allgemeine Ziel der Critical Infrastructure Protection (CIP) lässt sich als Gesamtheit aller Interessen der Stakeholder einer kritischen Infrastruktur umschreiben. Folglich müssen die zu schützenden Anliegen der Stakeholder betrachtet und zusammengefasst werden. Die Stakeholder lassen sich grob in drei Gruppen kategorisieren:

doi: 10.2313/NET-2015-03-1_09

Den Regulierer der Infrastruktur, den Infrastrukturbetreiber und den Infrastrukturnutzer.

Der Staat hat meistens die Rolle des Regulierers, wobei hier hierarchische Ebenen unterschieden werden müssen (in Deutschland zum Beispiel stehen EU-Regelungen vor Bundesregelungen vor Länderregelungen). Der Regulierer ist für das Verfassen von Vorschriften im Hinblick auf den Betrieb und auch den Schutz kritischer Infrastrukturen verantwortlich. Ziel seines Handelns ist ein kontrollierter Ablauf der Infrastruktur. Dadurch kann die Gesellschaft, in dessen Rahmen der Regulierer handelt, versorgt werden.

Ein Infrastrukturbetreiber besitzt eine Vielzahl an Möglichkeiten sich in einem Sektor am Betrieb einer Infrastruktur zu beteiligen. Verkörpert wird die Funktion überwiegend durch miteinander konkurrierenden Unternehmen. Als oberste Priorität setzen sie die Gewinnerzielung. Dieses Ziel kann nur in Verbindung mit einem erfolgreich geführten Betrieb erreicht werden, welcher auf der Zufriedenheit der Kunden basiert.

Der Nutzer (bzw. Kunde) kann je nach Sichtweise ein weiterer Infrastrukturbetreiber oder Regulierer sein. In der Prozesskette eines Infrastruktursystems stellt er jedoch den Endverbraucher dar. Sein Interesse liegt an einer verlässlichen Versorgung der nachgefragten Menge an Gütern und/ oder Diensten einer kritischen Infrastruktur.

Zusammenfassend lässt sich somit feststellen, dass sich die Interessen der drei Stakeholdergruppen grob in ein Ziel vereinen lassen. Sobald die Versorgung des Infrastrukturnutzers mit dem nachgefragten Gut oder der benötigten Dienstleistung gesichert ist, können die grundlegenden Anliegen aller Beteiligten mitberücksichtigt werden.

Im Folgenden werden nun die Ursachen untersucht, welche die reibungslose Versorgung, in Form von Störungen oder Ausfällen, unterbrechen kann, und deren Auswirkungen in Bezug auf das Energieversorgungssystem abgeschätzt.

4.1 Gefahren für kritische Infrastrukturen

Gefahren für kritische Infrastrukturen lassen sich dem Ursprung nach in drei Klassen einordnen[14]. Eine grobe Abschätzung des Ausmaßes der jeweiligen Ursachen wird in Bezug auf die drei Ebenen des Energieversorgungssystems getroffen (vgl. Kapitel 3).

4.1.1 Naturereignisse

Zu Naturgefahren gehören Extremwetterereignisse (Stürme, Hochwasser, Hitzewellen etc.), Waldbrände, seismische Ereignisse, Epidemien und kosmische Ereignisse.

Zwar sind das Ausmaß und die Eintrittswahrscheinlichkeit dieser Gefahren abhängig von der geographischen Lage, sie sind jedoch für die CIP der Energiesysteme von großer Bedeutung. Vor allem im Bereich der physischen Ebene befinden sich viele Elemente oberirdisch (Stromleitungen, Masten etc.) und sind somit vor allem für Extremwetterereignisse extrem anfällig.

Im November 2005 zeigte das ‚Münsterländer Schneechaos'[15] zum Beispiel, welchen Schaden Naturereignisse anrichten können, in diesem Fall starker Schneefall. 250.000 Menschen waren bis zu vier Tagen von der Stromversorgung abgeschnitten, da unter anderem 50 Strommasten unter der Last der Schnee- und Eisschichten zusammenbrachen.

4.1.2 Failure (technisch/ menschlich)

Systemversagen, Unfälle und Havarien, Fahrlässigkeit und organisatorisches Versagen sind der Kategorie technischer bzw. menschlicher Failure zuzuordnen.

Auf Grund der steigenden Vernetzung und daraus resultierenden Komplexität von Stromnetzen nehmen auch in diesem Gefahrenbereich die Risiken für den Betrieb zu[16]. Am Beispiel des deutschen Stromnetzes und dessen Integration auf europäischer Ebene entstehen kontinuierlich neue Herausforderungen im Hinblick auf die organisatorischen Fähigkeiten der Betreiber im Bezug auf deren Koordinations- und Kooperationsfertigkeiten. Dies führt vor allem auf der obersten Ebene (‚Entscheidungsebene') zu einem erhöhten Risiko.

Dieses Risiko wurde am 4. November 2006 zur Realität, als es auf Grund von mangelnden Sicherheitsmaßnahmen zu einem größeren Stromausfall[17] in Europa kam. Ungefähr 15 Millionen Menschen waren bis zu eineinhalb Stunden ohne Strom. Auslöser war die planmäßige Abschaltung einer von E.ON betriebenen Hochspannungsleitung für die Ausschiffung eines Kreuzfahrtschiffes. Durch die Abschaltung kam es zu einer Überlastung einer Verbindungsleitung, welche durch ihre automatische Abschaltung kaskardenartige Ausfälle über ganz Europa hinweg provozierte.

Neben den verwaltungstechnischen Herausforderungen, ist die beschleunigte technologische Entwicklung unter anderem Auslöser für eine Vielzahl an Neuerungen. Dies führt zu ungleichen Aktualisierungen von Soft- oder Hardwarekomponenten, was wiederum zu einer Inkompatibilität untereinander führen kann [14]. Hinzu kommt das Risiko, dass das Personal nicht laufend für die neuen Entwicklungen geschult wird bzw. werden kann und die Elemente folglich fehlerhaft bedient werden könnten.

4.1.3 Kriminalität

Sobald man Failure technischer oder menschlicher Art böswillig ausnutzt, kommt man in den Bereich der Kriminalität. Terrorismus, Sabotage, sonstige Kriminalität und (Bürger-)Kriege sind alles in allem schädigende Handlungen, die durch einen Menschen vorsätzlich ausgeführt werden.

Die zunehmende Interaktion zwischen IKT und Stromnetzen provoziert verstärkt Cyberattacken auf Energieversorgungssysteme. 53 Prozent aller Cyberattacken[18] sind auf den Energiesektor gerichtet. Eine Störung oder gar ein Ausfall der Energieversorgung würde weitreichende Konsequenzen auch in anderen Sektoren kritischer Infrastrukturen mit sich ziehen. Der Schutz vor virtuellen Anschlägen stellt folglich einen wichtigen Aspekt der nationalen CIP von Stromnetzen dar[19].

Das Bundesamt für Sicherheit in der Informationstechnik[20] identifizierte im Rahmen von Cyberangriffen die unberechtigte Nutzung von Fernwartungszugängen als eine der wichtigsten Bedrohungen. Wartungszugänge stellen die Schnittstelle eines IKT-Systems nach außen dar, sind aber jedoch heutzutage noch nicht ausreichend abgesichert.

Im Bereich der Stromnetze ist vor allem die Schnittstelle zwischen der physischen Ebene und den Kontrollzentren oftmals nicht ausreichend durch IT-Sicherheitsmaßnahmen geschützt. Die durch SCADA Systeme gesteuerte Schnittstelle stellt somit eine beliebte Angriffsfläche im Rahmen von Cyberattacken[21] dar. Wie E. Bompard et al.[12] schildern, sind die von SCADA Systemen verwendeten Kommunikationsprotokolle nicht durch Authen-

tifizierungs- oder Integritätsmechanismen geschützt. Dadurch ist es einem Angreifer zum Beispiel möglich Malware über Wechseldatenträger und Hardware einzuschleusen und auf die Datenflüsse zwischen den SCADA Systemen und den HMIs zuzugreifen. Damit kann er letztere mit irreführenden Informationen versorgen, um weitere Angriffe auf das Kontrollnetzwerk zum Beispiel zu verheimlichen.

Dieses Vorgehen und dessen enormen Auswirkungen rückten insbesondere durch den Wurm ‚Stuxnet'[22] an das Licht.

‚Stuxnet' adressierte ausschließlich Prozesssteuerungsrechner, auf denen die SCADA-Software ‚WinCC' von Siemens verwendet wurde. Sobald ein Angreifer über die Funktionen des SCADA Servers einer Prozesssteuerungsanlage verfügt, kann er nicht nur nicht autorisierte Befehle ausführen, sondern auch Datenkorruptionen durchführen oder das System anhalten.

4.2 Schutzmaßnahmen

Um Angriffen beziehungsweise Gefahren entgegenzuwirken, ist eine mögliche Herangehensweise durch drei Ansätze das Risiko einer Störung bzw. eines Ausfalls zu minimieren. Die Frage nach der Sicherheit eines Systems ist oberflächlich betrachtet die Voraussetzung für die Verlässlichkeit eines Systems und wird nicht getrennt untersucht. Im Folgenden werden die drei Prinzipien anhand von beispielhaften Lösungsansätzen für die analysierten Ursachen veranschaulicht.

4.2.1 Vermeidung der unmittelbaren Angriffswirkung

Das Prinzip der Vermeidung der unmittelbaren Wirkung eines schädigenden Ereignisses beläuft sich auf den, teilweise präventiven, Schutz der Komponenten einer Infrastruktur. Am Beispiel des Energieversorgungssystems kann Ursachenvermeidung unter anderem physisch betrieben werden, zum Beispiel durch die Errichtung von Mauern gegen Hochwasser oder durch Blitzableiter an Stromleitungen. Dadurch kann vor allem Naturgefahren entgegengewirkt werden.

Um insbesondere menschliches Versagen zu vermeiden ist das ‚Vier-Augen'-Prinzip ein verlässlicher Ansatz. Zu beachten ist, dass die über die zweite Person ausgeführte Kontrolle ernstgenommen wird. Wenn dies gewährleistet werden kann, kann die Korrektheit organisatorischer Entscheidungen im Bereich der Koordinations- und Kooperationsplanung zunehmend verstärkt werden. Jedoch muss man bedenken, dass durch den zunehmenden Vernetzungs- und Technologisierungsgrad der Stromnetze die Komplexität auf ein Maß steigt, welches vom menschlichen Auffassungsvermögen nicht mehr vollständig erfasst werden kann. Folglich ist die zunehmende Unterstützung des ‚Vier-Augen-Prinzips' durch IT-Systeme unerlässlich.

Schutzmaßnahmen können und müssen auch virtuell umgesetzt werden. Im Rahmen der Kriminalitätsbekämpfung hat das Thema Cyber-Sicherheit aktuelle Brisanz. Eine Studie zeigt[21], dass heutzutage nur 17 Prozent der befragten Unternehmen ausreichende IT-Sicherheitsmechanismen implementiert haben. Wie bereits im Unterkapitel 4.1.3 erörtert, liegt eine besonders große Schwachstelle der Cyber-Ebene im Bereich der SCADA Kommunikationsprotokolle. In den letzten Jahren wurden einige Protokolle vorgestellt, welche die Anforderungen an ein sicheres SCADA System erfüllen. Einerseits muss die Integrität der übermittelten Sensor- wie auch Anweisungsdaten gewährleistet werden. Andererseits sind die Authentizität der Kommunikationspartner, wie auch die Vertraulichkeit der Serverdaten grundlegende Anforderungen. Die Arbeitsgruppe 15 des Technischen Komitees 57 der Internationalen Elektrotechnischen Kommission (IEC) veröffentlichte Standards für die Cyber-Security in Energieversorgungssystemen. Weitere Informationen zu den publizierten Maßnahmen in [23].

4.2.2 Redundanz und Dezentralisierung

Die Verteilung der ausfallfähigen Elemente lässt sich durch zwei verschiedene Ansätze realisieren: der Redundanz einerseits und der Dezentralisierung andererseits.

Im Falle von Naturgefahren spielt die Redundanz eine wichtige Rolle. Ausschlaggebend hierfür ist das (n-1)-Kriterium. Dies besagt[24], dass im Falle des Ausfalls eines der n Versorgungswege die Versorgung ungestört fortgesetzt werden kann ohne dass dabei andere Elemente unzulässig belastet werden. Jedoch ist auch hier die steigende Vernetzung Grund für die Annahme, dass sobald zwei oder mehr Elemente gleichzeitig ausfallen, das (n-1)-Kriterium nicht mehr greift[25] und komplexere Prinzipien bei der Stromnetzplanung angewendet werden müssen.

Übergangselemente, zum Beispiel Haushaltsanschlüsse, Transformatoren sowie entsprechende Elemente im IT-Netzwerk, stellen in Stromversorgungseinrichtungen sogenannte Single Points of Failure (SPOF)[26] dar. Diese können im Falle eines Ausfalls zu erheblichen Störungen des Versorgungssystems führen. Durch redundante Geräte kann jedoch eine unterbrechungsfreie Versorgung gewährleistet werden[27]. Das redundante Gerät sollte hierbei zudem auf eine andere Art hergestellt worden sein, um die Wahrscheinlichkeit eines Ausfalls beider Geräte zu minimieren. In dem Falle eines Ausfalls des einen Gerätes kann durch die parallele Schaltung über Entkoppeldioden das andere Element einspringen ohne durch das defekte Gerät auch beschädigt zu werden.

Je dezentraler ein Energienetz aufgebaut ist, desto stabiler ist es nicht nur gegen Extremwetterereignisse, sondern im besonderen Maße gegen terroristische Angriffe. Beim Ausfall eines Kraftwerkes kann die Versorgung durch ein dezentralisiertes System trotzdem gesichert werden, da die Übertragungsnetze nicht nur von dem einen angegriffenen Kraftwerk abhängen[28]. Das Gesamtsystem wird somit robuster.

4.2.3 Unabhängigkeit

Unabhängigkeit von der Stromversorgung über das herkömmliche Verteilnetz ist insbesondere in Bereichen, wie dem des Gesundheitswesens, wichtig. Für alle drei Gefahrenbereiche gibt es einen einstimmigen Ansatz als Schutzmaßnahme.

Unabhängige Stromversorgungssysteme (USV) sorgen dafür, dass stromabhängige Geräte weiterlaufen[29] und somit der Betrieb der Infrastruktur nicht gestört wird. Zusätzlich sollen USV in der Lage sein, kurzzeitige Unter- und Überspannungen abzufangen. Damit sind Krankenhäuser zum Beispiel in der Lage für einige Zeit ohne das öffentliche Stromnetz weiter zu arbeiten.

Das Zukunftskonzept ‚Smart Grid' stellt eine weitere Möglichkeit dar im Falle eines Totalausfalls ‚intelligent' mit den vorhandenen Ressourcen auszukommen. Verbraucher sind durch ‚Smart Grids' in der Lage als sogenannte ‚Prosumer' zu agieren. Das heißt Strom, welchen ein Verbraucher (‚Consumer') zum Beispiel durch eigene Photovoltaik-Anlagen produziert, kann er je nach

Marktsituation selbst verbrauchen. Im Falle eines Ausfalls kann somit ein ‚Prosumer' seinen eigenen Strom verbrauchen und ist somit im gewissen Maße nicht von der öffentlichen Stromversorgung abhängig.

5. FAZIT

Kritische Infrastrukturen umfassen die Gesamtheit aller lebensnotwendigen Systeme und sind durch den Staat zu schützen. Insbesondere das Energieversorgungssystem ist auf Grund der hohen Vernetzung der verschiedenen Elemente untereinander auf eine funktionierende Interaktion aller beteiligten Akteure, physisch wie auch virtuell, angewiesen.

Die Verletzlichkeit kritischer Infrastrukturen und im besonderen Maße die des Stromnetzes, ist auf Grund der wachsenden Interdependenzen und Technologisierung in den letzten Jahren stark gestiegen. Der Schutz kritischer Infrastrukturen ist zu einer gesamtgesellschaftlichen Aufgabe herangewachsen, die nicht nur durch technische Maßnahmen ausgeführt werden kann. Mit Rücksicht auf die Durchdringung aller Bereiche durch IKT sollte in Zukunft vor allem Wert auf die Stärkung der Cyber-Sicherheit gelegt werden, zum Beispiel durch sichere Kommunikationsprotokolle. Dies sollte nicht nur durch staatliche Initiativen unterstützt, sondern vor allem unternehmensintern umgesetzt werden.

6. REFERENCES

[1] *The Report of the President's Commission on Critical Infrastructure Protection* (Oct. 1997). http://fas.org/sgp/library/pccip.pdf (16.09.2014).

[2] OECD: Protection of 'Critical Infrastructure' and the role of investment policies relating to national securities (May 2008). http://www.oecd.org/investment/investment-policy/40700392.pdf (16.09.2014).

[3] Bundesamt für Sicherheit in der Informationstechnik. https://www.bsi.bund.de/DE/Themen/KritischeInfrastrukturen/kritischeinfrastrukturen_node.html (20.09.2014).

[4] Department of Homeland Security: *National Infrastructure Protection Plan*. http://www.dhs.gov/xlibrary/assets/nipp_consolidated_snapshot.pdf (18.09.2014)

[5] Bundesamt für Bevölkerungsschutz und Katastrophenhilfe: *Sektoren- und Brancheneinteilung Kritischer Infrastrukturen*.http://www.bbk.bund.de/SharedDocs/Downloads/BBK/DE/Downloads/Kritis/neue_Sektoreneinteilung.pdf?__blob=publicationFile (17.09.2014).

[6] Cavelty, M./ Suter, M.: *The Art of CIIP Strategy: Tacking Stock of Content and Processes*. Erschienen in: Critical Infrastructure Protection – Information Infrastructure Models, Analysis, and Defense. Springer Verlag, 2012.

[7] Deutscher Bundestag: *Unterrichtung durch die Bundesregierung - Rahmenprogramm der Bundesregierung „Forschung für die zivile Sicherheit (2012 bis 2017)"* (Jan. 2012). http://dip21.bundestag.de/dip21/btd/17/085/1708500.pdf (22.10.2014).

[8] Negenborn, R./ Lukszo, Z./ Hellendoorn, H. (Hrsg.): *Intelligent Infrastructures*. Springer Verlag, 2010.

[9] Institute for Energy and Transport (IET), Joint Research Center. http://ses.jrc.ec.europa.eu/non-experts-0 (18.09.2014).

[10] Fraunhofer ESK: *Smart Grid Communications 2020* (Nov. 2011).http://www.esk.fraunhofer.de/content/dam/esk/de/documents/SmartGrid_Studie_final-web.pdf (20.09.2014).

[11] Bundesregierung: *Anteil Erneuerbarer Energien wächst weiter* (Jan. 2014). http://www.bundesregierung.de/Content/DE/Artikel/2014/01/2014-01-13-bdew-energiebilanz-2013.html (22.10.2014).

[12] Bompard, E. et al.: *Cyber Vulnerability in Power Systems Operation and Control.* Erschienen in: Critical Infrastructure Protection – Information Infrastructure Models, Analysis, and Defense. Springer Verlag, 2012.

[13] Tranchita, C. et al.: *ICT and Power Systems: An Integrated Approach.* Erschienen in: Securing Electricity Supply in the Cyber Age. Springer Verlag, 2010.

[14] Bundesministerium des Innern: *Nationale Strategie zum Schutz Kritischer Infrastrukturen (KRITIS-Strategie)*. http://www.bmi.bund.de/cae/servlet/contentblob/598730/publicationFile/34416/kritis.pdf (17.09.2014)

[15] DWD, Deutschländer, T./ Wichura, B.: *Klimastatusbericht 2005: Das Mümsterländer Schneechaos am 1. Adventswochenende 2005* (Nov. 2005). http://www.dwd.de/bvbw/generator/DWDWWW/Content/Oeffentlichkeit/KU/KU2/KU22/klimastatusbericht/einzelne__berichte/ksb2005__pdf/15__2005,templateId=raw,property=publicationFile.pdf/15_2005.pdfT. (20.10.2014).

[16] Bundesministerium des Innern: *Schutz Kritischer Infrastrukturen – Risiko- und Krisenmanagement*. http://www.bmi.bund.de/SharedDocs/Downloads/DE/Broschueren/2008/Leitfaden_Schutz_kritischer_Infrastrukturen.pdf?__blob=publicationFile (17.09.2014).

[17] Bundesnetzagentur: *Bericht über die Systemstörung im deutschen und europäischen Verbundsystem am 4. November 2006* (Feb. 2007). http://www.bundesnetzagentur.de/SharedDocs/Downloads/DE/Sachgebiete/Energie/Unternehmen_Institutionen/Versorgungssicherheit/Berichte_Fallanalysen/Bericht_9.pdf?__blob=publicationFile&v=1 (20.10.2014).

[18] Softpedia, Kovacs, E.: *ICS-CERT Warns of Brute-Force Attacks Against Critical Infrastructure Control Systems* (Jun. 2013). http://news.softpedia.com/news/ICS-CERT-Warns-of-Brute-Force-Attacks-Against-Critical-Infrastructure-Control-Systems-364266.shtml (20.10.2014).

[19] Bundesministerium des Innern: *Schutz kritischer Infrastrukturen – Basisschutzkonzept* (Aug. 2005). http://www.bmi.bund.de/SharedDocs/Downloads/DE/Broschueren/2005/Basisschutzkonzept_kritische_Infrastrukturen.pdf?__blob=publicationFile (17.09.2014).

[20] Bundesamt für Sicherheit in der Informationstechnik: *Angriffsmethoden* (Apr. 2012). http://allianz-fuer-cybersicherheit.de/ACS/DE/_downloads/angriffsmethoden/statistiken/ (22.10.2014).

[21] Cyber Risk Network, Ayers, E.: *Critical Infrastructure cyber risk scenarios not science fiction* (Jul. 2014).

http://www.cyberrisknetwork.com/2014/07/18/critical-infrastructure-cyber-risk/ (20.09.2014).

[22] Bundesamt für Sicherheit in der Informationstechnik: *Die Lage der IT-Sicherheit in Deutschland 2011* (Mai 2011). http://www.bsi.bund.de/SharedDocs/Downloads/DE/ (22.10.2014).

[23] International Electrotechnical Commission (IEC), Cleveland, F.: IEC TC 57 WG 15: IEC62351 Security Standards for the Power System Information Infrastructure. http://iectc57.ucaiug.org/wg15public/ (20.09.2014).

[24] Forschungsforum Öffentliche Sicherheit, Birkmann, J. et al.: *State of the Art der Forschung zur Verwundbarkeit Kritischer Infrastrukturen am Beispiel Strom/ Stromausfall.* Freie Universität Berlin, 2010.

[25] *Grundsätze für die Planung des deutschen Übertragungsnetzes* (Mär. 2012). http://www.50hertz.com/Portals/3/Content/Dokumente/Anschluss-Zugang/Verteiler/Planungsgrundsaetze-120330.pdf (20.10.2014).

[26] Jüllig, R.: *Analyse zur IT-Sicherheit in Energieversorgungssystemen* (Mai 2013). http://www.f07.fh-koeln.de/imperia/md/content/personen/waffenschmidt_eberhard/abschlussarbeiten_auschreibungen/juellig_it_sicherheit_forschungsarbeit2013.pdf (18.09.2014).

[27] Energie und Technik – Lexikon. http://energie-und-technik.de/ (20.09.2014).

[28] B.KWK, Golbach, A.: *Fakten und Thesen zur Dezentralisierung der Stromerzeugung* (Jul. 2004). http://bkwk.de/ (17.09.2014)

[29] Elektronik Kompendium. http://www.elektronik-kompendium.de/sites/grd/0812171.htm (20.09.2014).

doi: 10.2313/NET-2015-03-1_09

Out-of-Band Network Management

Felix Emmert
Betreuer: Oliver Gasser
Seminar Innovative Internet-Technologien und Mobilkommunikation WS2014
Lehrstuhl Netzarchitekturen und Netzdienste
Fakultät für Informatik, Technische Universität München
Email: felix.emmert@tum.de

ABSTRACT

Out-of-band network management is becoming more and more popular amongst high tech companies since high availability of network services is becoming more and more important. Especially servers with built in out-of-band management capabilities are growing in numbers as the increasing demand in bandwidth forces companies providing web services to outsource their networks to colocation centers that often lack the possibility for physical access. This article describes different benefits for network administrators, especially for administrators of out-of-band management devices for servers. Despite of the benefits this article mainly focuses on security issues surrounding out-of-band management devices for servers, analyzing the firmware of a Dell iDRAC 7 which is the latest out-of-band management device for Dell's rack servers. It shows what privileges potential attackers may gain by compromising out-of-band management devices. Finally some practical advise for system administrators on how to secure their systems against attacks is given.

Keywords

out-of-band, network management, IPMI, BMC, iDRAC

1. INTRODUCTION

Today it has become more and more popular for small and medium sized businesses to no longer host their web services in-house, but to rent computing power from specialized providers. The providers offer the needed infrastructure for storing, computing and delivering huge amounts of data in their colocation centers to many different customers. As a consequence of outsourcing computing needs to providers, it is in most cases no longer possible for system administrators to physically access their companies' servers.

Many systems like servers or networking hardware offer remote in-band management solutions. In-band management is accessible as long as the system is running but it cannot fully satisfy the need of physical access since it lacks efficient strategies to recover the system in case of emergencies like misconfigured networks or boot failures.

Out-of-band management aims to reduce the weak spots of in-band management by providing remote management functionalities independent of the system's operating state. In many cases this is achieved by independent sub-systems connected to the system they are managing (further called the "main system"). Many out-of-band management systems for servers have their own data storage (mainly flash storage) containing an own operating system as well as dedicated power supply and Ethernet ports. That way it is, for instance, possible to remotely power the main system on or off, access the system's keyboard, view the display's output and monitor the system's hardware even if the main system fails to boot.

While out-of-band management systems provide some major benefits for system administrators, they can also serve as backdoors for attackers. After gaining access it is possible to run various attacks like eavesdropping, compromising the main system or even using the out-of-band management system itself for future attacks (like as a botnet client). Once an attacker manages to modify the operating system running on an out-of-band management device, it can be very difficult to detect the intrusion and even more difficult to remove the threat since the malware survives a complete reinstallation of the main system. Malware residing on out-of-band management systems can actually carry over from one owner of the system to another, especially in the event of rented systems like servers in colocation centers.

This article mainly focuses on out-of-band management systems used to manage servers. It is showing their capabilities and security concerns surrounding them. It demonstrates some security issues analyzing Dell's latest iDRAC 7 firmware and gives advise on how to secure existing out-of-band server management devices.

2. RELATED WORK

Parts of this article are based on the work of Anthony J. Bonkoski, Russ Bielawski and J. Alex Halderman [1]. These authors analyze the security of IPMI (Intelligent Platform Management Interface), which is the industry standard for server oriented out-of-band management devices [9]. While they analyze Supermicro's implementation of an IPMI based out-of-band management system for servers they find many security issues similar to those found in Dell's iDRAC 7 which is analyzed in this article. Bonkoski, Bielawski and Halderman found an exploit inside the login system for the web interface of Supermicro's devices. This exploit is caused by the use of the insecure "strcpy" function. Bonkoski et al. developed a proof-of-concept buffer overflow attack to show the risks of the found exploit. Additionally they uncovered shell injection vulnerabilities which allow any user of Supermicro devices to execute system commands and even to run own code by downloading it onto the out-of-band manage-

doi: 10.2313/NET-2015-03-1_10

ment system using "wget" which is a standard GNU/Linux system tool for downloading web content. They found that at least 41,545 devices may be affected by these exploits thus being under immediate threat.

Recently Andrei Costin et al. published an article [2] which analyzes a set of 32,356 different embedded firmwares using a self designed automated system. While these firmwares are from all kinds of different devices the authors report similar security issues like the issues found in this article including the extraction of SSL certificates together with the according unencrypted private keys of about 35,000 devices connected to the Internet. They also found many hard-coded login credentials used for telnet, system or web logins. Additionally they discovered other backdoors like unsecure daemons, exploitable web interfaces and authorized SSH keys which can be used for remote connections.

3. OUT-OF-BAND MANAGEMENT TYPES

Out-of-band management systems are common in more sophisticated network hardware like servers. They provide system administrators with possibilities to manage their systems even in the event the main system is not running or otherwise unavailable.

Many routers, switches and hubs made by Cisco come with out-of-band management systems accessible via a "Network Management Module" connected to serial console ports on the devices [3] [4] or directly via a special ethernet ports on newer hardware [5]. That way it is possible to recover devices or entire networks that are no longer accessible via in-band management. The remote management capabilities can be extended to be accessible via backup networks or even wireless (GSM) by using aftermarket hardware [6].

In case of servers without built-in out-of-band management systems one of the most basic solutions for partial out-of-band access is the use of a KVM over IP device like the Peppercon LARA [7]. KVM means "Keyboard, Video, Mouse". KVM over IP devices enable their user to remotely view the system's display output and forward keyboard and mouse input to the system over a TCP/IP network. That way it is possible to interact with the system while it is still booting to change BIOS parameters or to fix boot issues. Some devices also include remote access to the system's hard reset and ON/OFF switch.

More dedicated servers offer built-in solutions like Dell's iDRAC, HP's iLO, Oracle's iLOM, and Lenovo's IMM. Those out-of-band management systems run on an embedded microcontroller called BMC (Baseboard Management Controller) which is integrated into the main system's hardware (either directly or via daughter card) [1]. The BMCs run their own operating system residing on dedicated data storage, mainly flash storage. In many cases the BMCs have access to the PCI bus, various I/O ports and sensors enabling it to fully control the server. Additionally the BMCs have their own network interface controllers (NICs) or at least access to one of the system's NICs via a "side-band" interface. The BMC may have an own power supply and/or a battery.

Figure 1 shows a basic setup of a server featuring an integrated out-of-band management device. It shows how the Dell iDRAC 7 device used for investigation in the making of this article has been connected to the Internet by a popular German colocation center operator.

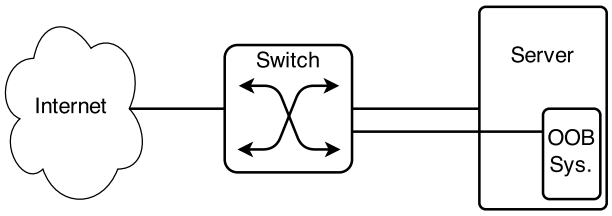


Figure 1: Basic setup of a server with an integrated out-of-band management device

4. SERVER MANAGEMENT

Most out-of-band management system implementations for servers are based on IPMI (Intelligent Platform Management Interface), which is the industry standard. In addition to IPMI many vendors offer additional user interfaces or functionalities. IPMI implementations commonly provide the following core functionalities.

4.1 Chassis Control

IPMI allows users to control the power state of the main system. It is possible to remotely power on or off the main system or perform a hard reset. Additionally vendors may allow to trigger a soft-shutdown (by emulating a fatal overtemperature) as well as pulse diagnostic interrupt or power cycling [9].

Additionally it is possible to perform other chassis operations like physically identifying itself (useful in colocation centers) or configuring power restore settings (like power on after AC power is restored).

4.2 System Provisioning

Many out-of-band management systems for servers include the possibility to provide the server with bootable media needed for installing a new operating system or booting into a live system for recovery. This can be done by simply selecting a physically connected drive to boot from (like USB flash drives) but it is also possible to remotely provide the system with image files (like iso) which are then connected to the server as virtual DVD drives [10]. Network boot via PXE may also be possible if supported by the server's BIOS.

4.3 KVM over IP

KVM over IP provides system administrators with remote access to the system's keyboard, video output and mouse. One major benefit of KVM over IP is the possibility to remotely debug system boot failures as no further software is needed to fully control the server.

Manufacturers often include KVM over IP abilities into their BMCs. To use built in KVM over IP, a vendor specific client application (most of the time Java based) is needed. This application connects to the BMC via TCP port (commonly port 5900).

doi: 10.2313/NET-2015-03-1_10

4.4 Watchdog Timer

A Watchdog Timer is a function designed to detect system malfunctions. To achieve that it is continuously decrementing a timer [9]. If this timer reaches the value of "0", specific actions like a system reset can be triggered. To prevent this behavior called "timeout", the system needs to continuously reset the timer to prove that it is still running. Apart from that it is possible to completely disable the Watchdog Timer.

Every BMC implementing IPMI must include a watchdog timer that is able to perform system resets [9]. Other possible timeout actions may be system power off or power cycle. Additionally, vendors may include pre timeout interrupt functionality triggering shortly before the actual timeout. This can be used to attempt a nonviolent system shutdown prior to hard resets.

4.5 Serial Over LAN

Serial Over LAN (SOL) is a feature allowing to remotely connect to the main system using the system's serial interface [1]. This can be used to remotely access the server's BIOS and bootloaders like Grub. Additionally some operating systems like GNU/Linux can be configured to accept serial console connections [8].

4.6 Web Interface

Many IPMI implementations operate an optional web interface on ports 80 and 443. The web interfaces enable users to check the system health, view and modify the BMC's settings and access various other features of the out-of-band management system like chassis control or system provisioning. Additionally many implementations allow users to download the application needed to access the KVM over IP feature directly from the BMC's web interface.

4.7 Command Line Interface

Despite web interfaces the most common way in accessing IPMI devices is via command line interface. IPMI specifies a protocol called IPMI over IP on UDP port 623 for this purpose. Additionally, many implementations host an optional SSH daemon on TCP port 22. The command line interface usually offers access to all BMC commands, settings and outputs the user is allowed to interact with.

5. SECURITY CONCERNS

Out-of-band management systems like IPMI devices are built to manage and restore the main system they are connected to. For being able to do so, they often need extensive control over their main system. While this is intended, it also brings huge security concerns along. An attacker that somehow gets access to an IPMI device will be able to misuse it for many different purposes. To make things worse, IPMI devices tend to have quite a large attack surface.

The following sections will shed light on some existing security issues, give an overview on possible threats imposed by compromised BMCs and advise on some best practice to harden IPMI enabled systems.

5.1 Dell iDRAC 7 Attack Surface

This section focuses on the attack surface of Dell's iDRAC 7, but it may also be true for other vendors' IPMI implementations.

A port scan on a Dell iDRAC 7 shows open TCP ports 80 and 443 for the internal appweb web server as well as TCP port 5900 used by the KVM over IP Java application. Additionally theres an SNMP agent on UDP port 161. Other services like an SSH daemon on TCP port 22, a TELNET daemon on TCP port 23, IPMI over IP on UDP port 623 and a VNC server on TCP port 5901 are deactivated by default on the investigated Dell iDRAC 7. All of these network services can be deactivated or configured to run on different ports with the exception of IPMI over IP whose port is fixed.

The internal web server of a Dell iDRAC 7 uses default SSL certificates which are not generated by the BMC upon initialization but instead shipped with the firmware, together with the appropriate unencrypted private keys. Although, it is possible to change the used SSL certificates, system administrators might fail to do so enabling attackers to decrypt the BMC's network traffic or set up phishing sites.

All remote management services do support basic password authentication which is enabled by default. This makes it possible to run attacks using known username/password combinations or brute force. If the system uses it's default SSL certificate for HTTPS traffic, it is also possible to acquire the login credentials by sniffing and decrypting the network traffic or phishing.

A severe weakness in the IPMI over IP protocol allowes for attackers to get the "HMAC" hash of the BMC's login credentials [12] [14]. This enables hackers to perform offline password cracking attacks. Since the BMC tells the attacker whether an username is valid or not without checking the password first, such an attack can be very efficient. This weakness exists since IPMI version 2.0.

While Dell's iDRACs do have default login credentials (root/calvin), the devices do encourage users to change them since firmware iDRAC 7 1.30.30 if the default login credentials are still in use [13]. On the investigated iDRAC 7 devices those credentials have been changed by the colocation center operator (Hetzner) prior to delivery of the system. Still there may be many outdated versions out there that don't warn their users. Additionally some administrators may not be aware of the existence of a BMC inside their servers hence not changing the credentials at all.

Other manufacturers also ship their devices with default login credentials. These credentials are shown in Table 1. The only manufacturer that uses random default passwords for shipping is HP.

5.2 Affected Systems

Having shown some weaknesses of IPMI devices, the next question would be the number of public reachable devices using insecure default settings. To get appropriate data, a network scanning tool is needed. Zmap is a powerful research tool capable of scanning the entire public accessible IPv4 range in less than an hour given enough bandwidth

Manufacturer	Default Username	Default Password
Dell	root	calvin
HP	Administrator	*random 8 char*
IBM	USERID	PASSW0RD
Supermicro	ADMIN	ADMIN
Fujitsu	admin	admin
Oracle/Sun	root	changeme
ASUS	admin	admin

Table 1: Common default login credentials of IPMI devices by manufacturer [14]

[15]. This analysis uses public available HTTPS scan data gathered by Zakir Durumeric et al. on the 29th of January 2014 [16].

The relevant part of the data consists of two tables, one containing all scanned HTTPS enabled IP addresses together with the SHA-1 fingerprint of the used certificate. The other table contains details about every certificate found in the scan. A first check against the SHA-1 fingerprint used in the most recent firmware of Dell's iDRAC 7 at the time of writing (1.57.57) shows 11,659 devices running on public IP addresses.

A deeper check was performed by gathering the SHA-1 fingerprints of every certificate containing the string "iDRAC" inside it's subject and matching against any of them. This results in a total of 46,490 devices on public IP addresses. Since this number contains iDRAC devices including older ones the subtraction of the two numbers shows that 34,831 devices may not be running on the latest iDRAC 7 firmware.

Due to administrators being able to change the certificate used by their devices this number only serves as a lower bound approximation of the total number of Dell iDRAC devices that are accessible via public IP addresses. However, the found devices may be very vulnerable because they do use the default certificate.

Further analysis could be done by accessing the home page of every IP hosting a web interface on well known HTTP or HTTPS ports by pattern matching against known contents like logos or headers.

5.3 Possibilities for Attackers

Attackers who gained access to an IPMI device can benefit from it in several ways. This section shows a selection of different possible use cases for Hackers.

5.3.1 Denial of Service

The easiest thing to do with a compromised BMC is to perform a Denial of Service (DoS) attack on the host system. Attackers can simply turn off the main system and prevent it from rebooting without having to modify any part of the software by changing the boot behavior. While this might not be the smartest of attacks as it is easy to detect the attack and restore the system, it could leave larger networks vulnerable to follow up attacks.

More stubborn attackers could modify the BMC to emulate false hardware faults or manipulate shared NICs to drop legit network traffic. Doing so would make it harder to detect the attack itself or its source (the BMC) possibly causing the system owners to believe that their server is damaged.

5.3.2 Eavesdrop

Once an attacker manages to compromise the BMC, the attacker could start eavesdropping without attracting attention. This can be achieved in various ways.

One method would be packet sniffing on the NIC. This could target the main system as well as other systems connected to the same network. Attackers could try to gain system passwords or capture various sessions. It would even be possible to perform man-in-the-middle attacks trying to spy on weakly encrypted network traffic.

Eavesdropping may also be done by using the KVM over IP features (since KVM over IP shows the server's display output) or by analyzing different system logs.

5.3.3 Take over the main system

One of the goals of many attacks may be getting control over the main system connected to the out-of-band management system. Since the BMC is designed to control the main system, this attack is not very difficult. An attacker could simply boot some live operating system and mount the harddrive as root. That way the attacker will be able to modify any part of the server's operating system to his needs.

Although encrypting the system drive may prevent this from happening, it would still be possible to modify the BMC's system and eavesdrop on the encryption password the next time an administrator enters it at a system reboot. Since system reboots can be enforced using the BMC's abilities attackers won't have to wait for an opportunity (which could take quite some time as it is common that servers are rarely rebooted at all).

Additionally it might also be possible to use the KVM features to gain access to the main operating system without the need of rebooting it.

5.3.4 Persistent rootkits

Attackers who manage to modify the operating system of the BMC may be able to install highly persistent rootkits due to the closed and independent nature of the BMC. Such malware will survive any action taken to clean the main system like reinstallation or even a complete replacement of all system drives as the BMC uses it's own storage. Additionally it may remain undetected for a long period of time and possibly carry over to new owners [1].

Rootkits could be further enhanced by being able to detect firmware updates or resets and modify the new firmware on the fly. Doing so would make it close to impossible for most administrators to ever get rid of the rootkit without replacing the whole out-of-band management device.

5.3.5 BMC botnets

Another scenario would be using the out-of-band management hardware itself for future attacks like in a botnet [1].

Botnets are large networks of infected systems which can be used for large scale attacks on single targets (like DDoS attacks) or for attacking huge amounts of different targets. Botnets can also serve as a source of computing power that can be used to do massive calculations like for example cracking passwords or generating crypto currency.

While it might not seem to be very promising to use BMCs for botnets due to their limited computing power, the lifespans of BMC botnets may be very long if combined with persistent rootkits. Other advantages of such botnets would be their huge network bandwidth and availability due to the fact that many servers featuring an out-of-band management system are located inside colocation centers or other facilities with massive network backbones.

Such botnets could come into existence in very short periods of time if attackers manage to remotely capture BMCs as they mostly run a very limited variety of different firmwares. Infected BMCs have already been reported [11] and it is just a small step to combine them into networks.

6. IDRAC 7 FIRMWARE ANALYSIS

This section will focus on a deeper analysis of Dell's iDRAC 7 firmware 1.57.57 which is the most recent firmware for the iDRAC 7 found in a Dell PowerEdge R720 at the time of writing. The firmware is available on Dell's public ftp server [17].

The downloaded .EXE file can be extracted using UnZip 6.0 [18]. The resulting files include the actual firmware image inside a folder named "payload". Examining the image file named "firmimg.d7" using Binwalk v2.1.0 [19] shows that it contains a Linux kernel followed by two "squashfs" filesystems. These filesystems can be mounted using the offset provided by Binwalk or extracted using Binwalk itself.

The first and bigger filesystem contains the Linux root filesystem, the other filesystem contains various default settings as well as installation scripts.

It is possible to detect modifications of the firmware since it is signed by an ASCII armored PGP signature at the end of the file.

Further inspection of the root filesystem shows that it contains the iDRAC 7 default SSL certificate together with it's unencrypted private key. This alone generates a huge risk as attackers may setup phishing sites with the exact same certificate as found on the original iDRAC web interface. So even if some people trust the default certificate of their iDRAC 7 believing it was generated upon first initialization instead of being shipped with the firmware, they also trust every other iDRAC 7 with a similar firmware and they trust the phishing sites. If done right, a phishing site can not be distinguished from a default iDRAC 7 web interface. Such a site may yield a lot of valid login credentials if propagated to administrators managing Dell iDRACs. Since the iDRAC 7 devices used for investigation always got a public IP address that was in the same /30 IPv4 subnet as the main server it may be quite easy to get a lot of correct email addresses of related system administrators.

User	Password as salted MD5 hash
root	1fY6DG6Hu$OpwCBE01ILIS1H/Lxq/7d0
user1	1nVOr80rB$HDAd6FRlG24k/WN4ZuYPC0

Table 2: Account names and password hashes found in the shadow file inside Dell's iDRAC 7 firmware 1.57.57

Additionally to having the ability of creating phishing websites using the default certificate attackers could also decrypt any captured network traffic between a Dell iDRAC 7 and it's operator using the default certificates. This may also provide them with login credentials or they could act as a man-in-the-middle by altering and re-encrypting the network traffic.

Besides that the system's shadow file (shown in Table 2) containing two shell enabled system accounts together with salted MD5 hashes of their passwords has been found. Since the password hash of the root account did not match with the default credentials (root/calvin) these may be the credentials for the underlying GNU/Linux system as opposed to the login credentials for IPMI access.

Dell's iDRAC 7 seems to be running on a Renesas SuperH H4 CPU. Since this architecture is not very common, it is hard to find any free decompiler that works. Instead the disassembly tools of the GNU toolchain for the Renesas SH7751R CPU [20] have been used to inspect parts of the web interface back-end.

Basic analysis of the web interface back-end which is inside the
iDRAC's cgi-bin shows that C's "strcpy" function well known for it's security issues [21] has been used within multiple parts, including the login function. This could potentially result in buffer overflow attacks. Since in-depth analysis of assembly code would go beyond the scope of this article, it has not been further analyzed. If a buffer overflow attack is possible that would mean that anybody could access a Dell iDRAC 7 that runs on a public IP. Moreover this can be an entry point for code injections rendering the system extremely vulnerable to more complex attacks like rootkits.

7. HARDENING IPMI DEVICES

Since IPMI devices tend to have security issues this section aims to give advice on how to harden IPMI devices. Following these suggestions will result in a lower attack surface presented to the public Internet.

Any operator of an IPMI device providing a web interface should install a custom SSL certificate. This certificate should not include information about the nature of the device to make it harder to identify. It is critical that the new certificate is not uploaded using a public network. Doing so would provide attackers with the new certificate since they can decrypt the network traffic which is encrypted with the default SSL certificate. Instead a private network should be used to transfer the certificate. In case of a Dell iDRAC 7 it is also possible to use SSH to do so if the SSH host keys identifying the system have already been securely exchanged. On the investigated Dell iDRAC 7 devices these keys have been

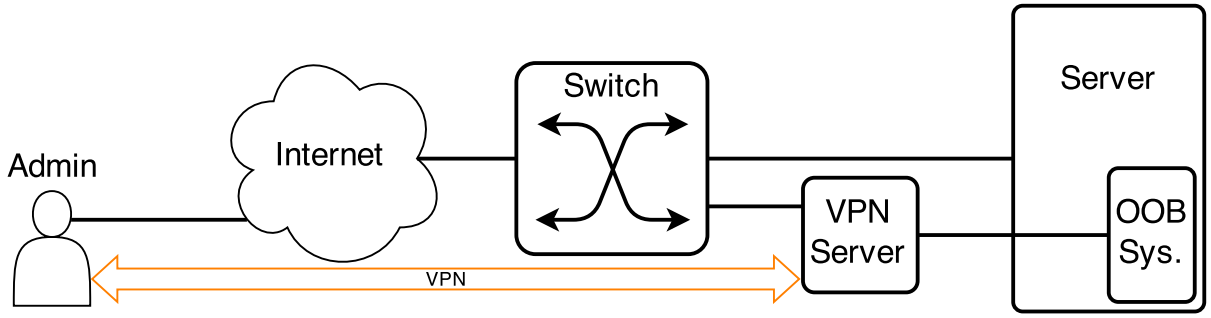


Figure 2: Out-of-band management device secured by a VPN

generated by the BMC upon first initialization.

Additionally, to changing the certificate IPMI devices should always operate inside a secure, closed network. Especially IPMI over IP should never be enabled outside of a secure network due to it's security weakness. In case of some IPMI implementations like a Dell iDRAC 7 it is possible to use VLAN tagging [22] to separate the device from the public Internet. Since remote access is needed in colocation center environments, secure VPNs like IPsec or SSH tunnels can be used. The gateways used to forward traffic to the closed network of IPMI devices have to be secured. Other possibly vulnerable parts of the internal network like web or mail servers should not have access to the gateways. In case high availability is crucial redundant gateways can be used.

Figure 2 shows an example setup of a server with an integrated BMC which is operating in a separated network. Yet it is still reachable by administrators through a VPN. Other BMCs can be connected to the VPN server by adding a network switch between the BMCs and the VPN server.

8. CONCLUSION

Out-of-band management devices do provide users with a lot of useful tools to manage their network devices especially if these devices reside inside colocation centers. Administrators can use them to install, supervise and recover servers they cannot physically interact with. While the benefits provided by such devices may be quite interesting it is important to consider their security flaws. Users of these devices have to be aware of the risks but it is the manufacturers responsibility to make their devices secure especially at times of high profits gained through industrial espionage.

This article showed some of the many features of IPMI devices together with some possible scenarios of what attackers can do with them. The firmware 1.57.57 of a Dell iDRAC 7 has been analyzed which is the most recent firmware at the time of writing. The results show some security issues encouraging administrators to take immediate actions. Finally some practical advice on how to lower the attack surface of IPMI devices has been given.

Administrators should never ignore their IPMI devices since they often run out of the box without any need to be enabled first, especially if the BMC is connected to the public Internet. This is even more important since at least 46,490 Dell iDRAC devices and potentially even more devices made by other manufacturers are running on public IPs. It seems to be common practice of colocation center operators to connect IPMI devices to the Internet after setting up new servers for customers.

9. REFERENCES

[1] Anthony J. Bonkoski, Russ Bielawski, and J. Alex Halderman: *Illuminating the Security Issues Surrounding Lights-Out Server Management*, In Proceedings of the 7th USENIX Workshop on Offensive Technologies (WOOT '13), August 2013

[2] Andrei Costin, Jonas Zaddach, Aurélien Francillon, and Davide Balzarotti, Eurecom: *A Large-Scale Analysis of the Security of Embedded Firmwares*, In Proceedings of the 23rd USENIX Security Symposium, August 2014

[3] Cisco: *FastHub 300 Series Installation and Configuration Guide*, chapter Out-of-Band Management, `http://www.cisco.com/c/en/us/td/docs/switches/lan/hubs/fhub316c_t/install_config/guide/fh300icg/rptroutb.pdf`

[4] Cisco: *FastHub 300 Series Hubs Network Mgmt Module Instal Note*, `http://www.cisco.com/c/en/us/td/docs/switches/lan/hubs/fhub316c_t/expansion_mods/install/notes/4089_01.html`

[5] Cisco: *Cisco ASA 5500 Series Configuration Guide using the CLI, 8.4 and 8.6*, chapter Configuring Interfaces, `http://www.cisco.com/c/en/us/td/docs/security/asa/asa84/configuration/guide/asa_84_cli_config/interface_start.html`

[6] Perle: *Console Servers for Out of Band Management of Cisco Routers, Switches and Firewalls*, `http://www.perle.com/supportfiles/cisco_tech_note.shtml`

[7] Daxten: *Peppercon LARA - KVM remote administration*, `http://www.daxten.com/uk/kvm-over-ip.html`

[8] ArchWiki: *Working with the serial console*, October 2014, `https://wiki.archlinux.org/index.php/working_with_the_serial_console`

doi: 10.2313/NET-2015-03-1_10

[9] Intel, Hewlett-Packard, NEC, and Dell: *Intelligent Platform Management Interface Specification Second Generation*, October 2013, http://www.intel.com/content/dam/www/public/us/en/documents/product-briefs/ipmi-second-gen-interface-spec-v2-rev1-1.pdf

[10] Paul Ferrill, ServerWatch: *Server Management Tools: A Closer Look at HP's iLO and Dell's iDRAC*, http://www.serverwatch.com/server-reviews/server-management-tools-comparison-a-closer-look-at-hps-ilo-and-dells-idrac.html

[11] Web Hosting Talk forum post: *SuperMicro IPMI Security*, October 2010, http://www.webhostingtalk.com/showthread.php?t=992082

[12] NIST, National Cyber Awareness System: *Vulnerability Summary for CVE-2013-4786*, October 2013, http://web.nvd.nist.gov/view/vuln/detail?vulnId=CVE-2013-4786

[13] Dell TechCenter: *iDRAC7 now supports Default Password Warning feature*, iDRAC7nowsupportsDefaultPasswordWarningfeature

[14] HD Moore, Metasploit: *A Penetration Tester's Guide to IPMI and BMCs*, https://community.rapid7.com/community/metasploit/blog/2013/07/02/a-penetration-testers-guide-to-ipmi

[15] Zakir Durumeric, Eric Wustrow, and J. Alex Halderman: *ZMap: Fast Internet-Wide Scanning and its Security Applications*, In Proceedings of the 22nd USENIX Security Symposium, August 2013

[16] Zakir Durumeric, James Kasten, Michael Bailey, and J. Alex Halderman: *HTTPS Ecosystem Scans*, https://scans.io/study/umich-https

[17] Dell US: *DELL iDRAC 1.57.57 Driver Details*, http://www.dell.com/support/home/us/en/19/drivers/DriversDetails?productCode=poweredge-r720&driverId=XH6FX

[18] Info-ZIP: *Info-ZIP's UnZip*, http://www.info-zip.org/UnZip.html

[19] Binwalk: *Firmware Analysis Tool*, http://binwalk.org/

[20] Renesas: *Linux & Open Source @ Renesas*, SH7751R Linux BSP, https://oss.renesas.com/modules/download/index.php?cid=52

[21] CERN Computer Security: *Common vulnerabilities guide for C programmers*, https://security.web.cern.ch/security/recommendations/en/codetools/c.shtml

[22] IEEE Standards for Local and metropolitan area networks: *Virtual Bridged Local Area Networks*, IEEE Std 802.1Q™, 2003 Edition

doi: 10.2313/NET-2015-03-1_10

Smart Buildings vs. Data Privacy Law

Michael Keil
Betreuer: Dr. Holger Kinkelin, Marcel von Maltitz
Seminar Innovative Internettechnologien und Mobilkommunikation WS 2014/2015
Lehrstuhl Netzarchitekturen und Netzdienste
Fakultät für Informatik, Technische Universität München
Email: Michaelkeil8590@aol.com

KURZFASSUNG

Um die Energieeffizienz von Gebäuden und Stromnetzen zu verbessern, wird eine große Menge an unterschiedlichen Daten, z.B. Energieverbräuche und Präsenz-Daten, gesammelt. Da diese Daten in manchen Situationen benutzt werden können um Personen eindeutig zu identifizieren, und aus ihnen andere Informationen wie Verhaltensmuster abzuleiten, treten Datenschutzgesetze in Kraft um vor Missbrauch persönlicher Daten zu schützen. Inwieweit Datenschutzgesetze auf das Sammeln von Daten in intelligenten Gebäuden zutreffen, welche Auswirkungen diese haben und welche Anforderungen erfüllt werden müssen um in keinen Konflikt mit den Datenschutzgesetzen zu geraten wird in dieser Ausarbeitung dargelegt.

Schlüsselworte

Datenschutz, Intelligente Gebäude, Smart Building

1. EINLEITUNG

Heutzutage legen viele Menschen immer größeren Wert auf die Erzeugung von erneuerbarer Energie. Um die erzeugte Energie so gut wie möglich nutzen zu können, ist es sehr wichtig neue Technologien und Systeme zu entwickeln und zu installieren die eine optimale Verteilung ermöglichen. Zusätzlich muss hierfür der Energieverbrauch gemessen werden und neue, bessere Technologien zum Speichern gewonnener Energie müssen erforscht werden, um den überschüssigen Strom, der zum Beispiel Nachts erzeugt wird, nicht zu verschwenden. Aus diesen Beweggründen sind viele Staaten dazu übergegangen Pläne für intelligente Stromnetze, sogenannte „Smart Grids“, zu entwickeln. Die Bundesnetzagentur spricht von einem „Smart Grid“ wenn ein konventionelles Elektrizitätsnetz mit Kommunikations-, Mess-, Steuer-, Regel-, und Automatisierungstechnik sowie IT-Komponenten aufgerüstet wird um den Netzzustand in „Echtzeit“ zu erfassen und Möglichkeiten zur Steuerung und Regelung des Netzes bestehen, um die Netzkapazität optimal nutzen zu können. [1] Während über die Datenschutzprobleme, die mit dem Erfassen von Energiedaten in „Smart Grids“ entstehen, in vielen Ländern diskutiert wurde, ist diese Problematik, in intelligenten Gebäuden, sogenannten Smart Buildings, bisher nur wenig beachtet worden. Auch ist diese Problematik bei intelligenten Gebäuden wesentlich schwerer zu erfassen, als bei der Erfassung des Energieverbrauches in intelligenten Stromnetzen. Während es bei Smart Grids offensichtlich ist, dass die Daten die in Haushalten erfasst werden geschützt werden müssen, da auch Haushalte erfasst werden in denen einzelne Personen oder Familien leben, ist die Problematik bei intelligenten Gebäuden weniger gut zu erkennen. Intelligente Gebäude zielen darauf ab Energieeffizienz, Komfort und Sicherheit im gesamten Gebäude zu gewährleisten, die Datenerfassung bezieht sich deshalb hauptsächlich auf Sensor- und Energiedaten welche in den meisten Fällen gesetzlich nicht geschützt werden müssen, jedoch gibt es auch Ausnahmen in denen die erfassten Daten unter den gesetzlichen Schutz fallen und ein gesetzeskonformer Umgang gewährleistet sein muss. Deshalb beschäftigt sich diese Arbeit damit, die Problematik des Datenschutzes in intelligenten Gebäuden darzulegen und aufzuzeigen welche Datenschutzgesetze eingehalten werden müssen und ob vorhandene Strategien zur Entwicklung datenschutzkonformer Software angewendet werden können, um die Vorgaben die hinsichtlich Datensicherheit und Datenschutz bestehen, erfüllen zu können.

In Kapitel 2 wird hierfür der Begriff „Smart Building“ definiert und es wird dargelegt welche Daten in diesen Gebäuden gesammelt werden. Kapitel 3 befasst sich dann mit den bestehenden Datenschutzgesetzen und der Fragestellung warum diese auf die gesammelten Daten intelligenter Gebäude Anwendung finden. In Kapitel 4 werden Hoepmanns Strategien dargelegt und es wird in Kapitel 5 untersucht ob diese auf die Problemstellung anwendbar sind.

2. SMART BUILDING GRUNDLAGEN

Der Auf- und Ausbau von „Smart Buildings“ ist ein wichtiges Thema, das auch in Zukunft noch weiterhin betrachtet werden wird.

2.1 Definition

Verschiedene Definition zum Begriff „Smart Building“ sind heutzutage vorhanden. der „Smart 2020“-Bericht liefert anhand von fünf Begriffen eine Definition zum Thema Informations- und Kommunikationstechnologien (ICT) in Gebäuden.

In diesem Bericht werden folgende fünf Worte verwendet um den Begriff „smart“ im Bezug auf Gebäude zu definieren.

Standardise: Der Umgang mit Informationen zum Energiekonsum und Emissionen in Systemen und Produkten der ICT soll standardisiert sein.

Monitor: Daten sollen in „Echtzeit“ überwacht und überprüft werden um die Energieeffizienz des Gebäudes zu erhöhen.

Account: Dem Konsumenten werden aufbereitete Daten, zu Themen wie Energieverbrauch und Emissionen, zur Verfügung gestellt um diesen in einer Verbesserung der Energieeffizienz einzubinden.

doi: 10.2313/NET-2015-03-1_11

Rethink: Durch die übermittelten Informationen soll der Konsument angestoßen werden sein Verhalten zu überdenken und bewusster mit Energie umzugehen.

Transform: Letztendlich soll der Konsument sein Energieverhalten verändern um die Energieeffizienz, an Stellen an denen eine Automatisierung nicht möglich ist, zu verbessern. [3][4]

Eine andere Definition liefert die Siemens AG. Diese ist der Meinung, dass nur Lösungen, welche die größte Synergie zwischen Energieeffizienz, Komfort und Sicherheit besitzen über längere Zeit bestehen bleiben werden. Lösungen die Gebäude in lebendige Organismen verwandeln, die vernetzt, intelligent, sensibel und anpassungsfähig sind. [5]

Intelligente Gebäude sind also Gebäude, die die Energieeffizienz verbessern, sich gegebenen Umständen anpassen, Wartungskosten reduzieren, indem nur notwendige Wartungen durchgeführt werden anstelle von Wartungen die nach einem vordefinierten Zeitplan stattfinden, und einen erhöhten Komfort liefern.

2.2 Erfassbare Daten

Um diese Ziele erreichen können werden in intelligenten Gebäuden eine Vielzahl von Daten gesammelt.

2.2.1 Temperatur

Temperatursensoren messen die vorherrschende Temperatur und leiten diese weiter zum Beispiel an eine Anzeigevorrichtung oder einen Datenspeicher. Ein Beispiel hierfür ist das Thermometer

2.2.2 Bewegungsdaten

Ein weiterer Sensor der heutzutage in den meisten größeren Gebäuden vorkommt ist der Bewegungssensor. Dieser erkennt die Bewegung eines Objektes oder Subjektes und kann dadurch Aktionen auslösen.

2.2.3 Stromverbrauch

Eine der wichtigsten Datenmessungen, um die Energieeffizienz zu erhöhen, ist das Aufzeichnen des Stromverbrauches. Je detaillierter man diesen misst und auswertet umso einfacher ist es eine Steigerung der Energieeffizienz zu erzielen.

2.2.4 Luftqualität und Luftfeuchtigkeit

Auch die Luft wird gemessen und analysiert. Neben der Qualität werden vor allem Rauchsensoren angebracht um auf eventuelle Gefahrensituationen reagieren zu können.

2.2.5 Statussensoren

Eine weitere Möglichkeit Ursachen für eventuelle unerwünschte Situationen zu finden ist das Messen von Statusinformationen. So wird zum Beispiel gemessen ob ein Fenster oder eine Tür geöffnet ist, in welchem Stock sich zurzeit der Aufzug befindet oder ob Leuchten angeschaltet sind oder nicht.

2.2.6 Kameraüberwachung

Um sicherheitsrelevante Bereiche abzudecken besteht auch die Möglichkeit Überwachungskameras anzubringen und somit optische Daten zu sammeln.

2.2.7 Sonstiges

Auch können weitere Sensoren angebracht sein um zu erkennen ob eine Glasscheibe intakt ist und wie hell es in einem Raum ist. Neben gebäudeinternen Daten werden auch gebäudeexterne Daten gesammelt zum Beispiel durch die Benutzung von Wetterstationen[6]

2.3 Datenverwendung

Diese Daten werden für unterschiedliche Aktionen verwendet. Um die Sicherheit von Personen im Gebäude und des Gebäudes selbst zu gewährleisten, werden in vielen Gebäuden sicherheitsrelevante Plätze Videoüberwacht um Bedrohungen rechtzeitig erkennen und beheben zu können. Neben Kameras dienen vor allem auch Glasbruchsensoren dazu ein unbefugtes Betreten von außerhalb zu erkennen, um somit Personen und Gegenstände im Gebäude zu schützen. Aber auch Gefahrensituationen innerhalb des intelligenten Gebäudes sollen mit Hilfe dieser Daten erkannt und gelöst werden. Beispiele dafür sind das Ausbrechen eines Feuers, oder das Auftreten gesundheitsgefährdender Gase in der Luft. Diese werden durch Rauchsensoren und Sensoren zur Überprüfung der Luftqualität erkannt und es können Gegenmaßnahmen eingeleitet werden, wie zum Beispiel die Räumung des Gebäudes und die Verständigung zuständiger Stellen.

Neben der Sicherheit ist der Komfort ein wichtiger Faktor intelligenter Gebäude. So helfen Temperatursensoren die Belüftungsanlagen zu steuern um eine angenehme Temperatur innerhalb des Gebäudes zu gewährleisten und Lichtsensoren steuern ein automatisches Herablassen des Sonnenschutzes um den Aufenthalt im Gebäude so komfortabel wie möglich zu gestalten.

Neben Komfort und Sicherheit spielt der Energieverbrauch eine wichtige Rolle. In dieses Thema fallen wie schon angesprochen das Thema Beleuchtung aber auch die Verwendung von Statussensoren hilft ein Gebäude effizienter zu machen. Ein schnelles Erkennen und mögliches automatisches Schließen eines offenen Fensters im Winter hilft dabei zum Beispiel. Um Gebäude jedoch noch Energieeffizienter gestalten zu können, zu hohen Energieverbrauch messen zu können und dazu beizutragen, dass nur der Strom verbraucht wird, der auch benötigt wird, ist es wichtig den Energieverbrauch von Geräten so genau wie möglich zu bestimmen.

Viele Sensoren helfen bei mehr als einem dieser Problem. So trägt der Bewegungssensor einerseits dazu bei den Stromverbauch des Gebäudes zu senken indem er die Lampensteuerung übernimmt, und Leuchten nur angehen wenn sie gebraucht werden, sondern erhöht auch gleichzeitig den Komfort innerhalb des Gebäudes, da das Betätigen von Lichtschaltern wegfällt.

Jedoch unterliegt das Sammeln und Verwenden von bestimmten Daten gesetzlichen Vorschriften. Dies führt dazu, dass ein gesetzeskonformes System entwickelt und eingesetzt werden muss um die Datensicherheit und den Datenschutz zu gewährleisten.

3. DATENERFASSUNG VS. DATENSCHUTZ

Während es einen internationalen Standard zum Thema Datenschutz gibt (ISO 29100 [7]) an dem man sich orientieren kann, ist es wichtig das Erfassen und Weiterverarbeiten von Daten so zu gestalten, dass es mit dem zutreffenden Gesetz kompatibel

doi: 10.2313/NET-2015-03-1_11

ist. In Deutschland gilt zum Beispiel das Bundesdatenschutzgesetz (BDSG) an das man sich halten muss. Im folgenden werden einige themenbetreffende Auszüge aus diesem Gesetz vorgestellt.

3.1 Bundesdatenschutzgesetz (BDSG)

Für intelligente Gebäude ist es sehr wichtig Daten so genau wie möglich zu sammeln um eine Optimierung der Energieeffizienz gewährleisten zu können. In manchen Fällen können jedoch Personen anhand eines Datensatzes identifiziert werden. Beispielsweise wenn der Energieverbrauch eines bestimmtes Computers gemessen wird, der nur von einer einzigen Person benutzt wird. Die so gesammelten Daten werden dann als personenbezogene Daten nach BDSG §3 bezeichnet.

Wodurch die komplette Datenerfassung den Paragraphen 9 des BDSG erfüllen muss. Wäre es möglich eine Zuordnung grundsätzlich zu verhindern, würde dem Bundesdatenschutzgesetz ebenfalls genüge getan.

(1) „Öffentliche und nicht-öffentliche Stellen, die selbst oder im Auftrag personenbezogene Daten erheben, verarbeiten oder nutzen, haben die technischen und organisatorischen Maßnahmen zu treffen, die erforderlich sind, um die Ausführung der Vorschriften dieses Gesetzes, insbesondere die in der Anlage zu diesem Gesetz genannten Anforderungen, zu gewährleisten.

(2) Erforderlich sind Maßnahmen nur, wenn ihr Aufwand in einem angemessenen Verhältnis zu dem angestrebten Schutzzweck steht.“ [8]

Da Daten in intelligenten Gebäuden automatisiert gesammelt werden, muss die Datenerfassung konform zu den Anlagen des §9 sein. Diese besagen, dass

1. eine Zutrittskontrolle stattfinden muss, um zu gewährleisten, dass unbefugte Personen keinen Zutritt zu den Räumlichkeiten haben, in denen die gesammelten Daten verarbeitet werden. Ein Beispiel hierfür ist ein Firmenausweis, der nur Befugten das Betreten der Datenverarbeitungsanlagen gestattet.

2. Es muss eine Zugangskontrolle stattfinden, die nur Befugten die Benutzung gewährleistet. Es dürfen also nur befugte Personen den Computer benutzen auf dem die Daten gespeichert und verarbeitet werden. Ein Beispiel für diese Zugangskontrolle ist eine Passwort-Kontrolle, sodass niemand anderes Zugang zu den Daten hat.

3. Es muss weiterhin gewährleistet sein, dass Befugte auch nur Zugriff auf die Daten haben, für die sie eine Zugriffsberechtigung haben und die personenbezogenen Daten bei der Nutzung, Verarbeitung und nach der Speicherung nicht unbefugt gelesen, kopiert, verändert oder entfernt werden können .

4. Personenbezogene Daten bei der elektronischen Übertragung, ihres Transports oder ihrer Speicherung auf Datenträger nicht unbefugt gelesen, kopiert, verändert oder entfernt werden können. Außerdem muss überprüft und festgestellt werden können an wen diese Daten übertragen werden dürfen.

5. Es muss eine Eingabekontrolle stattfinden. Das heißt es muss nachvollziehbar sein wer die personenbezogenen Daten gesammelt, verändert oder entfernt hat.

6. personenbezogene Daten dürfen nur so verarbeitet werden wie der Auftrag es zulässt. Das heißt zum Beispiel wenn eine Zeitung Email-Adressen sammelt um Neuigkeiten an die jeweiligen Personen übermitteln zu können, darf die Email-Adresse auch nur dafür verwendet werden.

7. Es muss gewährleistet werden, dass personenbezogene Daten gegen zufällige Zerstörung oder Verlust geschützt sind.

8. Außerdem muss das System die Möglichkeit zur Verfügung stellen das Daten, die für unterschiedliche Zwecke gesammelt wurden auch getrennt voneinander verarbeitet werden können.

Außerdem ist es für die Punkte zwei bis vier besonders wichtig, dass keine veralteten Verschlüsselungsverfahren verwendet werden, da dies ein Verstoß gegen Satz 2 darstellen würde. Sollte außerdem ein Arbeitsverhältnis vorliegen, sodass der Arbeitgeber die Daten des Arbeitnehmers sammelt oder verarbeitet tritt zusätzlich noch BDSG § 32 in Kraft der die Datenerhebung, -verarbeitung und -benutzung für Zwecke des Beschäftigungsverhältnisses regelt.[8]

3.2 Problem der Datenerfassung

Wie beschrieben gelten diese Datenschutzparagraphen nur für personenbezogene Daten, wäre es unter keinen Umständen oder nur mit einem unverhältnismäßigen großen Aufwand an Zeit, Kosten und Arbeitskraft möglich mit Hilfe der gesammelten Daten eine Person zu identifizieren, würde die Datenschutzgesetze nicht in Kraft treten. Außerdem würde die in BDSG §3a verlangte Anonymisierung wegfallen, die verlangt das personenbezogene Daten soweit wie möglich anonymisiert werden um eine Identifikation zu erschweren. [8]

Um eine genauere Erklärung aufzuzeigen, betrachten wir das Problem anhand der Energiemessung eines Computers (Abbildung 1). In der Abbildung sieht man das Starten des

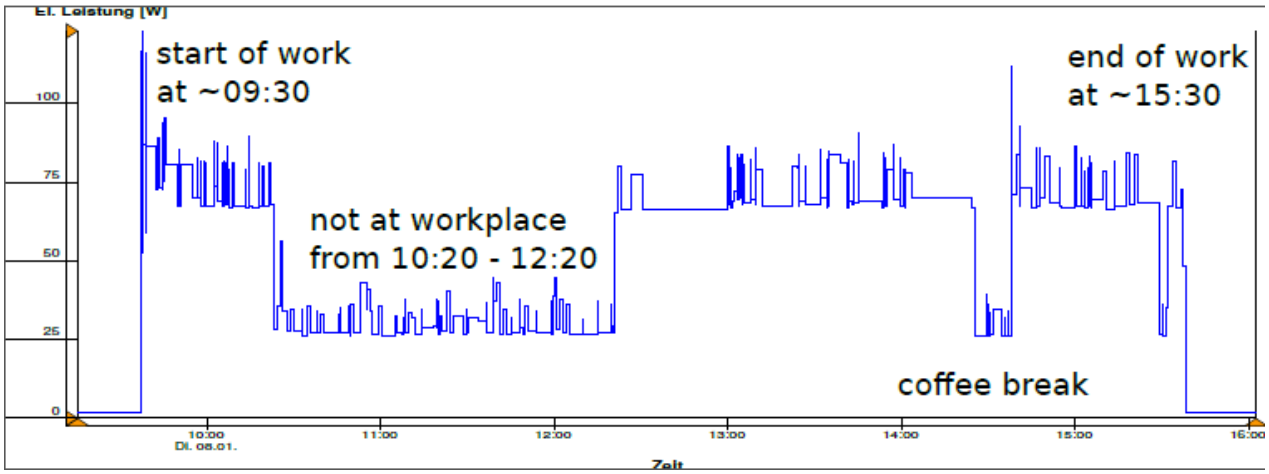


Abbildung 1. Energiemessung eines Computers[11]

doi: 10.2313/NET-2015-03-1_11

Computers um ungefähr 9.30 Uhr. Von ca. 10.20 Uhr bis 12.20 Uhr benötigt der Computer weniger Energie. Dies deutet darauf hin, dass der Computer zu dieser Zeit nicht benutzt wurde. Um 14.30 Uhr ist der Computer wieder nicht in Benutzung. Wenn man diese Daten jetzt mit der Energiemessung der Kaffeemaschine des Büros vergleicht, wäre es möglich Rückschlüsse führen zu können, ob der Benutzer zu dieser Zeit einen Kaffee getrunken hat. Auch wäre es möglich diese Zeit mit den Meldungen von Türen zu vergleichen um eine mögliche Route nachzubilden.

Neben Energiemessungen können aber auch andere Daten zu bestimmten Zeitpunkten oder in unterschiedlichen Kombinationen dazu benutzt werden um ein Individuum zu identifizieren. So können Bewegungsdaten von Bewegungsmeldern, oder Temperaturdaten des Büros dazu genutzt werden einen Bewegungsablauf zu bestimmen. Besonders einfach wäre dies bei Messungen in einem Büro, in dem nur eine Person arbeitet. Auch die Daten einer Zutrittskontrolle sollten auf keinem Fall in Verbindung mit anderen Daten genutzt werden dürfen. Selbst wenn die Daten der Zutrittskontrolle anonymisiert sind, sobald es möglich ist Festzustellen ob nur eine Person ein Gebäude betreten hat, stellt diese Information in Verbindung mit vielen anderen Daten eine Grundlage zu einer Identifizierung da. So können Sensordaten von Türen zum Beispiel dazu genutzt werden ein Bewegungsprofil zu erstellen. Dieses Profil kann mit Informationen darüber welche Türen geöffnet wurden zu einer Identifizierung führen.

Aber auch sensible Temperatur- und Luftqualitätssensoren könnten in kleinen Räumen Schwankungen aufweisen, die es gestatten könnten, den Aufenthalt einer Person in diesem Raum zu bestimmen. Diese Information könnte dann wiederum in Kombination mit anderen Daten wie zum Beispiel einer automatischen Lampensteuerung auf den Gängen dazu benutzt werden Bewegungsprofile anzulegen, welche wie beschrieben zu einer Identifizierung führen könnte.

In solchen Fällen sind diese Daten dann personenbezogene Daten und müssen, um sie nutzen zu können, entweder anonymisiert oder geschützt werden und den Anlagen des §9 des BDSG entsprechen. Sollte eine Anonymisierung nicht möglich sein dürfen die geschützten personenbezogene Daten nur nach Einwilligung der jeweiligen Person benutzt werden. Außerdem hat die jeweilige Person nach Charta der Grundrechte der europäischen Union das Recht zu erfahren welche Daten gesammelt wurden [8][9].

Nachdem das Problem der Datensammlung von „Smart Buildings“ erkannt wurde, stellt sich die Frage ob es möglich ist bestehenden Softwareentwicklungsstrategien zu folgen um eine datenschutzkonforme Erfassung und Verarbeitung für intelligente Gebäude zu entwickeln.

4. DATENSCHUTZSTRATEGIEN VON HOEPMANN

Neben der ISO 29100 gibt es auch verschiedene weitere Strategien an denen man sich orientieren kann um den Datenschutz einzuhalten. Darunter fallen auch die von Hoepmann vorgestellten Strategien, die als Unterstützung dienen sollen um datenschutzkonforme Software zu entwickeln. Diese orientieren sich an den europäischen Gesetzen und Richtlinien und erhöhen, bei einer Umsetzung dieser Strategien, die Wahrscheinlichkeit das die entwickelte Software gesetzeskonform in Europa ist.

4.1 Beschreibung und Erklärung der Datenschutzstrategien

Hoepmann entwickelte folgende acht Strategien um Datenschutzprobleme auszuschließen. Die Strategien sind Minimieren (Minimise), Verbergen (Hide), Trennen (Separate), Aggregieren (Aggregate), Informieren (Inform), Kontrollieren (Control), Durchsetzen (Enforce) und Demonstrieren (Demonstrate)

4.1.1 Minimieren

Unter „Minimieren“ versteht Hoepmann, dass

„die Menge der personenbezogener Daten die verarbeitet wird auf die geringste mögliche Menge begrenzt ist.“[10]

Das heißt, dass man, wenn möglich, keine personenbezogenen Daten sammelt und verarbeitet. Sollte es jedoch nicht möglich sein ohne diese Daten, dass gewünschte Ergebnis zu erzielen soll die Menge der persönlichen Daten so gering wie möglich gehalten werden.

4.1.2 Verbergen

Die zweite Strategie ist das „Verbergen“.

„Jede Art von persönlichen Daten, und ihre Beziehungen zueinander, sollen von jeglicher Betrachtungsmöglichkeit aus verborgen werden.“ [10]

Der Hintergedanke dieser Entscheidung ist, dass es nicht möglich ist die persönlichen Daten zu missbrauchen, wenn man keine Einsicht in diese hat. In welchem Ausmaße diese Strategie umgesetzt werden sollte ist jedoch situationsabhängig. Nehmen wir an ein Arbeitgeber braucht einige persönliche Daten eines Angestellten um diesem Angestellten in einer Angelegenheit zu helfen. Wenn der Arbeitgeber das Einverständnis des Angestellten hat, seine Daten für diesen Zweck zu benutzen, müssten in diesem Fall die persönlichen Daten nur vor Dritten verborgen werden.

4.1.3 Trennen

Bei der Strategie „Trennen“ geht es darum, dass persönliche Daten einzelner Personen getrennt voneinander und von verschiedenen Abteilungen bearbeitet werden. Dies soll verhindern, dass ein komplettes Profil einer Person angelegt werden kann, aufgrund der Tatsache, dass man nur einzelne Daten dieser Person besitzt. Auch die Speicherung persönlicher Daten sollte in verschiedenen, nicht verbundenen Plätzen stattfinden, um zu verhindern, dass Unbefugte die Möglichkeit besitzen diese Daten zusammenzufügen.

4.1.4 Aggregieren

Nachdem in der Strategie „Trennen“, die persönlichen Daten einzelner Personen getrennt wurden, soll in der Strategie „Aggregieren“ ein Zusammenfügen ähnlicher Daten stattfinden.

„Persönliche Daten sollen hier in größtmöglichen Gruppierungen und mit dem kleinstmöglichen Detailreichtum, in der die Daten noch nutzbar sind, verarbeitet werden.“ [10]

Hinter dieser Entscheidung steht die Idee, dass gruppierte Datensätze an Detailreichtum verlieren, da beispielsweise nur Durchschnittswerte der Daten verwendet werden, wodurch Spitzen eines einzelnen Datensatzes verloren gehen, und somit

von diesen Daten weniger auf einzelne Personen geschlossen werden kann. Jedoch kann man hier nicht nur auf die Größe der Gruppierungen achten, da der Datensatz nach dem Gruppieren immer noch eine sinnvolle Verarbeitung zulassen muss. Wenn man beispielsweise den Energieverbrauch von Computern mit dem Energieverbrauch von Leuchten gruppiert, könnte man aus dieser Gruppierung wahrscheinlich keine sinnvollen Daten mehr extrahieren.

4.1.5 Informieren

Während die ersten vier Strategien besonders darauf abzielen wie die Daten verarbeitet werden sollen, decken „Informieren“, „Kontrollieren“, „Durchsetzen“ und „Demonstrieren“ das Selbstbestimmungsrecht über die eigenen persönlichen Daten ab.

Die Strategie „Informieren“ ist dafür der erste Schritt. Personen müssen nach §33 BDSG informiert werden wenn ihre personenbezogenen Daten gespeichert oder verwendet werden.

Sie müssen informiert werden, welche Daten wie verwendet werden und welches Ziel hinter der Verwendung steht. Außerdem müssen sie auf Anfrage darüber informiert werden wie ihre persönlichen Daten geschützt werden und ob diese Daten mit Dritten geteilt werden. Auch muss Betroffenen auf Nachfrage mitgeteilt werden, welche Daten zum jeweiligen Zeitpunkt noch gespeichert sind.

4.1.6 Kontrollieren

Da wie gesagt jeder das Recht hat selbst über seine personenbezogenen Daten zu bestimmen, müssen auch Möglichkeiten zur Verfügung gestellt werden, dass Betroffene eigene Daten betrachten, aktualisieren und wenn erwünscht auch löschen können. Diese Strategie steht im besonders engen Zusammenhang mit der „Informieren“-Strategie, da es weder sinnvoll ist Betroffene über alles zu informieren wenn sie jedoch keinerlei Möglichkeit besitzen diese Daten zu kontrollieren. Aber auch jegliche Möglichkeit der Kontrolle ist nicht sinnvoll, wenn Betroffene in nicht ausreichender Weise über die Verwendung ihrer persönlichen Daten informiert werden.

4.1.7 Durchsetzen

Durchsetzen bezeichnet die Strategie, dass

„Datenschutzstrategien, die die rechtlichen Anforderungen erfüllen, bestehen und durchgesetzt werden sollen.“[10]

Diese Strategie ist sehr wichtig um eine gesetzeskonforme Software zu entwickeln. Besonders wichtig ist, dass jeder Punkt der Strategie umgesetzt wird. Es darf beispielsweise nicht vorkommen, dass man eine Methodik entwickelt hat die eine rechtliche korrekte Ausführung zulässt, diese jedoch nicht umsetzt. Auch müssen verwendete Techniken und Maschinen auf dem aktuellen Wissensstand sein um Verstöße gegen das Datenschutzgesetz abzufangen. Ein Beispiel hierfür ist das Verwenden von aktuellen Verschlüsselungsverfahren anstelle von alten Verfahren, welche möglicherweise günstiger in der Verwendung sind jedoch nicht mehr als sicher angesehen werden.

4.1.8 Demonstrieren

Die letzte der acht Strategien, ist die Strategie des „Demonstrierens“. Diese verlangt, dass man demonstrieren kann, dass die entwickelte Software allen Datenschutzvorgaben gerecht wird. Datenschutzbeauftragten soll es mit der Umsetzung dieser Strategie möglich sein, eine effektive Implementierung der Gesetze zu zeigen und bei Problemen die Reichweite der möglichen Datenschutzprobleme zu identifizieren.

4.2 Initiale Anwendungsgebiete

Ursprünglich entwickelte Hoepmann seine Strategien als Design Strategien die Softwareentwickler unterstützen sollten, datenschutzkonforme Software zu entwickeln. Jedoch ergibt sich durch die Aufgliederung der Strategien und die Umsetzung der in Europa vorherrschenden Datenschutzgesetze die Möglichkeit diese Strategien auch zum Evaluieren bereits existierender Software Design Patterns und Software zu benutzen.

Abbildung 2 zeigt eine Tabelle, die zeigt, welche Strategien auf bestehende gesetzliche Richtlinien angewendet werden können. So wird zum Beispiel eine gesetzliche geforderte Minimierung der verwendeten personenbezogenen Daten in den Strategien „Minimieren“, „Verbergen“ und „Aggregieren“ großflächig abgedeckt, während eine Benachrichtigung von Betroffenen bei einer Verletzung des Datenschutzgesetzes, also wenn zum Beispiel Unbefugte sich Zugang zu diesen Daten geschaffen haben, in der Strategie „Informieren“ abgedeckt ist.

	Aufgabenbegrenzung	Datenminimierung	Datenqualität	Transparenz	Selbstbestimmung der Betroffenen	Adäquater Schutz	Benachrichtigung bei Verletzung des Datenschutzes	Nachweisbare Einhaltung der Datenschutzgesetze
Minimieren	o	x						
Verbergen		x				o		
Trennen	o					o		
Aggregieren	o	x						
Informieren				x			x	
Kontrollieren			o	x	x			
Durchsetzen	x		x		x	x		o
Demonstrieren				o				x

Legende:
"x": große Abdeckung
"o": geringfügige Abdeckung

Abbildung 2. Abbildung der Strategien auf gesetzliche Prinzipien

5. ANWENDBARKEIT VON HOEPMANNS STRATEGIEN AUF DAS BESTEHENDE DATENSCHUTZPROBLEM

Nachdem Hoepmanns Strategien dargelegt wurden, stellt sich die Frage ob sich diese Strategien, die eine sinnvolle Unterstützung zur Entwicklung von datenschutzkonformer Software darstellen, auf das Datenschutzproblem, das sich bei der Datensammlung in intelligenten Gebäuden ergibt, anwenden lassen.

doi: 10.2313/NET-2015-03-1_11

5.1 Strategieanwendung

Da sich bei Anwendbarkeit und Einhaltung aller acht Strategien ein System ergibt, dass den Datenschutzgesetzen in Europa entspricht, wäre es wünschenswert wenn diese Strategien nicht nur auf Softwaresysteme Anwendung finden würden. Deshalb wird hier die Anwendbarkeit jeder einzelnen Strategie auf das vorliegende Problem überprüft, um mit Hilfe dieser Strategien ein gesetzeskonformes System für dieses Problem entwickeln zu können.

5.1.1 Minimieren

Die „Minimieren"-Strategie verlangt, dass die Erhebung personenbezogener Daten so weit wie möglich minimiert wird. Da in intelligenten Gebäuden nicht gezielt personenbezogene Daten gesammelt werden, sondern die gesammelten Daten personenbezogene Daten sind, da man in Einzelfällen diese Daten bestimmten Personen zuordnen kann, ist eine Minimierung im Grunde schon umgesetzt. Eine weitere Minimierung wäre es nur wenn man das Sammeln von Datensätzen die einer Person zugeordnet werden können unterbindet, was jedoch nicht umsetzbar ist weil sich diese Situation verändern. Zum Beispiel an einer Kaffemaschine in einem Büro holen sich normalerweise viele Leute einen Kaffee, wodurch dieser Datensatz alleine keine Zuordnung zu einer einzelnen Person zulässt. Während der Urlaubszeit arbeitet jetzt nur eine Person in diesem Büro, wodurch eine Zuordnung möglich ist und das verarbeitende System jetzt das Bundesdatenschutzgesetz erfüllen muss. Eine Minimierung auf die Benutzung keiner personenbezogenen Daten ist also nicht möglich. Inwiefern eine Umsetzung der „Minimieren"-Strategie umsetzbar ist, ist jedoch auch abhängig vom Ziel der Datennutzung. Wenn dieses Ziel trotz Minimierung des Datensatzes erfüllbar ist, kann und sollte eine Reduzierung des Datensatzes vorgenommen werden. Da in Smart Buildings jedoch eine Vielzahl unterschiedlicher Daten gesammelt werden, besteht in den meisten Fällen die Möglichkeit den erfassten Datensatz weiter zu reduzieren. Auch sollte mit dem Hinzufügen und Entfernen von Diensten die Größe des Datensatzes jedes mal neu bestimmt werden, um einen minimalen Datensatz zu gewährleisten.

5.1.2 Verbergen

Da in intelligenten Gebäude eine automatisierte Sammlung stattfindet, wäre es möglich die Strategie „Verbergen" soweit umzusetzen, dass niemand Zugriff auf personenbezogene Daten hat, sondern eine automatische Verarbeitungssoftware die Daten soweit anonymisiert, dass kein Rückschluss auf eine Person mehr möglich ist. Somit hätten weitere Bearbeiter nur Zugriff auf eine Datensammlung, die keine personenbezogene Daten mehr enthält. Sollte es jedoch nicht möglich sein bestimme personenbezogenen Daten zu anonymisieren oder zu gruppieren, sollte ein Whitelistingansatz eingeführt werden, sodass nur einzelne Personen Zugriff zu bestimmten Datensätzen haben. Auch sollte hier die Strategie „Trennen" umgesetzt werden, damit einzelne Personen nicht Zugriff zu verschiedenen Datensätzen haben, was die Wahrscheinlichkeit einer Zuordnung zu einer Person erhöht. Es wäre also am sichersten wenn für jeden Datensatz der nicht automatisiert verarbeitet werden kann, nur eine einzige Person Zugriff zu diesen Daten erhält, die keine Genehmigung hat andere Datensätze einzusehen und zu bearbeiten.

5.1.3 Trennen

Da in Smart Buildings eine Vielzahl an verschiedenen Diensten installiert ist und angeboten wird, werden für deren Umsetzung auch eine riesige Menge Daten erfasst. Diese Daten sollten aufgrund der „Trennen"-Strategie separiert gespeichert und behandelt werden. Da einige Mehrwertdienste zur Umsetzung jedoch mehr als einen Datensatz benötigen, könnte hier wiederum ein Whitelistingansatz realisiert werden, sodass nur bestimmte Mehrwertdienste Zugriff auf mehrere Datensätze erhalten und dann auch nur auf die Datensätze die zur Erfüllung des Dienstes benötigt werden. Auch kann überlegt werden, ob man sich für jeden Mehrwertdienst einzeln die Einwilligung zur Verwendung der Nutzerdaten beschafft. Auf diese Weise kann jeder Nutzer selbst entscheiden welche Daten von welchen Diensten verwendet werden dürfen, wodurch der Nutzer die Kontrolle über die Trennung seiner Daten erhält.

5.1.4 Aggregieren

Auch ist es in Smart Buildings möglich Daten auf unterschiedliche Art und Weisen zu aggregieren. Die erste Möglichkeit ist eine räumliche Aggregation. Hier können zum Beispiel Daten aus Abteilung oder Geschossen zusammengefasst werden, um die Informationstiefe zu verschleiern. Es wäre möglich den Stromverbrauch aller Computer einer Abteilung zu aggregieren, so werden Benutzungsdaten einzelner Computer verschleiert, jedoch ist es weiterhin möglich den Stromverbrauch während unterschiedlicher Tageszeiten zu erfassen.

Neben einer räumlichen Aggregation kann auch eine zeitliche Aggregation stattfinden. Man gruppiert Datensätze also über einen ganzen Tag oder ganze Wochen. Auf diese Weise kann kein Benutzungsschema erstellt werden, was zu einer Zuordnung von Personen führen könnte, jedoch ist es beispielswiese weiterhin möglich den Stromverbrauch der Kaffeemaschine zu erfassen. Wenn man jetzt weiterhin erfasst wie viele Kaffee ausgegeben wurden und wie oft die Kaffemaschine sich selbst gereinigt hat, kann über längere Zeit trotz Aggregation festgestellt werden in welchen Phasen die Kaffemaschine wie viel Strom benötigt und ob die Anschaffung einer anderen Kaffeemaschine Einsparungen mit sich bringen könnte oder nicht.

5.1.5 Informieren

Eine Umsetzung der Strategie „Informieren" ist jedoch nur schwer möglich da man hierfür alle Datensätze, die eine Identifizierung zulassen, erkennen müsste und man sich auch mit der jeweiligen Zuordnung sicher sein müsste, um eine Weitergabe an Dritte ausschließen zu können. Die Information weiterzugeben welche Daten in einem intelligenten Gebäude gesammelt werden ist jedoch einfach zu erzielen, dies könnte mithilfe von öffentlicher Displays bewerkstelligt werden. Auch könnten Informationen auf einzelne Räume beschränkt werden. So könnte zum Beispiel in einem intelligenten Besprechungszimmer nur angezeigt werden, welche Daten in diesem Zimmer gesammelt werden. Auf diese Weise wären Benutzer darüber informiert welche Daten auf welche Weise erfasst werden.

5.1.6 Kontrollieren

Wie auch bei der Strategie „Informieren" ist das Anwenden von „Kontrollieren" für bestimmte Daten nur schwer erzielbar, da erst eine eindeutige Identifizierung dieser Datensätze stattfinden müsste um den betroffenen Personen die Möglichkeit zu geben diese Daten zu kontrollieren. Jedoch bietet das öffentliche

Informieren darüber welche Daten gesammelt werden dem Benutzer die Möglichkeit sich in vielen Fällen bewusst dagegen zu entscheiden. Würden Benutzer eines Besprechungszimmer darüber informiert werden welche Daten in diesem Besprechungszimmer erfasst werden, könnten Benutzer, die damit nicht einverstanden sind, das Besprechungszimmer verlassen oder die Besprechung könnte an einen anderen Ort verlegt werden. Auch könnte man diese Informationen vorab weiterleiten, sodass ein passender Ort gefunden werden kann.

5.1.7 Durchsetzen

Nachdem ein Lösungsansatz gefunden wurde um den Datenschutz in intelligenten Gebäuden einzuhalten, ist es möglich eine Methodik zu entwickeln den Datenschutz einzuhalten und es besteht auch die Möglichkeit diese Methodik durchzusetzen. Beispielsweise könnten man Beauftragte anstellen, die in regelmäßigen Abständen überprüfen ob nicht genehmigte Sensoren angebracht wurden beziehungsweise ob alle genehmigten Sensoren noch funktionsfähig sind. Damit kann ausgeschlossen werden, dass nicht genehmigte Daten von neuen Sensoren erfasst wurden. Außerdem stellt man so sicher, dass Benutzer korrekt informiert werden, welche Daten erfasst werden. Sollte ein Sensor defekt sein und man würde den Nutzer informieren, dass diese Daten gesammelt werden, würde man dem Nutzer ohne das Erfassen dieser Daten die Kontrolle entziehen. Auch könnten diese Beauftragten überprüfen ob Zutritts- und Zugangskontrollen richtig umgesetzt wurden und die Datenverarbeitungsanlagen gesetzeskonform gehandhabt werden.

5.1.8 Demonstrieren

Auch das „Demonstrieren“ könnte zu Problemen führen, da Lösungsansätze zur Einhaltung der Datenschutzgesetze in intelligenten Gebäuden mit hoher Wahrscheinlichkeit nicht verständlich für jeden sein werden und somit eine Demonstration und Erklärung schwer möglich wird. Außerdem müssten Umsetzungen, die ein anschauliches und verständliches Erklären ermöglichen, bei jeder kleinen Änderung des Systems aktualisiert werden und es müsste erneut erfasst werden ob diese neue Umsetzung allgemein verständlich ist. Auch ist in den meisten Fällen die Reichweite von Problemen schwer abzuschätzen, da selbst ein kleines Problem riesige Auswirkungen haben kann.

Im Allgemeinen lässt sich sagen, dass eine Abdeckung aller acht Strategien zwar umsetzbar ist, es jedoch noch einige Probleme zu überwinden gibt um dies zu erzielen.

5.2 Abdeckung vorhandener Problemlösungsansätze

An der Technischen Universität München wurde ein Ansatz für ein Energiemanagementsystem (EMS) entwickelt, das die Datenschutzgesetze, welche bei der Sammlung von Energiedaten Anwendung finden, einhalten soll. Dieses EMS deckt zurzeit einen Großteil von Hoepmanns Strategien ab. An Punkten wie „Informieren“, und „Demonstrieren“ muss in Zukunft jedoch noch gearbeitet werden. [11]

6. ZUSAMMENFASSUNG

Zusammenfassend ist zu sagen, dass das Thema Datenschutz in der heutigen Zeit ein sehr wichtiges und schwer umzusetzendes Thema darstellt. Auch wenn die Einführung von „Smart Grids“ und der Ausbau und Bau von „Smart Buildings“ für die Zukunft meiner Meinung nach der einzig richtige Weg ist, um eine energieeffizienteres Netzwerk aufzubauen. Es müssen jedoch noch einige Probleme überwunden werden. Ganz besonders das Informieren von Personen, deren Daten gesammelt wurden, stellt zur Zeit noch ein Problem da. Ein gute Möglichkeit auf den laufenden zu bleiben was das Thema Datenschutz bei „Smart Grids“ und „Smart Buildings“ betrifft, ist den „Smart Grid“ Aufbau in Großbritannien zu verfolgen, wo zur Zeit Vorschläge zu diesem Thema eingeholt werden.[2]

7. REFERENCES

[1] Bundesnetzagentur, Eckpunktepapier - "Smart Grid" und "Smart Market", Dezember 2011, Online verfügbar unter http://www.bundesnetzagentur.de/DE/Sachgebiete/ElektrizitaetundGas/Unternehmen_Institutionen/NetzentwicklungundSmartGrid/SmartGrid_SmartMarket/smartgrid_smartmarket-node.html, Letzter Aufruf am 2014/12/17

[2] Department of Energy & Climate Change, Smart Grid Vision and Routemap, February 2014

[3] The Climate Group. Smart Report 2020. Enabling the low carbon economy in the information age, 2008 Online verfügbar unter http://www.smart2020.org.

[4] Dominik Blunshy, Smart Buildings Einsatz von ICT in Gebäuden zur Steigerung der Energieeffizienz., 2010

[5] Siemens, Smart buildings - the future of building technology, 2010, Video online verfügbar unter https://www.youtube.com/watch?v=gCuPx9shWT0

[6] Sean Barker, Aditya Mishra, David Irwin, Emmanuel Cecchet, and Prashant Shenoy, Jeannie Albrecht, Smart*: An Open Data Set and Tools for Enabling Research in Sustainable Homes, SustKDD 2012

[7] ISO/IEC 29100. Information technology – Security techniques – Privacy framework. Technical report, ISO JTC 1/SC 27.

[8] Gola/Schomerus BDSG, Bundesdatenschutzgesetz 10. Auflage 2010

[9] Charta der Grundrechte der europäischen Union, (2010/C 83/02)

[10] Jaap-Henk Hoepman. Privacy design strategies, 2012. Online verfügbar unter http://arxiv.org/abs/1210.6621; Letzter Aufruf am 2014/12/18.

[11] Holger Kinkelin, Marcel von Maltitz, Benedikt Peter, Cornelia Kappler, Heiko Niedermayer, Georg Carle. Privacy Preserving Energy Management, 2014, Online verfügbar unter http://idem-project.de/downloads.php; Letzter Aufruf am 2014/12/18

Moving Target Defense

Bettina Noglik
Betreuer: Schlamp Johann
Seminar Innovative Internettechnologien und Mobilkommunikation
Lehrstuhl Netzarchitekturen und Netzdienste
Fakultät für Informatik, Technische Universität München
Email: noglik@in.tum.de

KURZFASSUNG
Das Paper gibt eine breit gefächerte Einführung in die Moving Target Defense (MTD). Sowohl die theoretischen Ansätze als auch die Bezüge zu übergeordneten Konzepten werden erklärt und Problemstellungen und Herausforderungen werden dargestellt. Drei Beispiele sollen daraufhin einen Einblick in die Anwendung in der Praxis geben.

Schlüsselworte
Netzwerksicherheit, Computersicherheit, Moving Target Defense

1. EINLEITUNG
Heutzutage haben die meisten Informationssysteme eine eher statische Konfiguration bzw. behalten eine bestimmte Konfiguration über eine relativ lange Zeitdauer. Das ermöglicht beispielsweise Angreifern Systeme in aller Ruhe auszukundschaften und dann passende Angriffe zu starten.

Das ist beispielsweise beim *Heartbleed*-Bug geschehen. Nachdem der Fehler im Programmcode der Open SSL-Bibliothek gefunden und sich die Information in Hackerkreisen verbreitet hatte, wurden im großen Stile vertrauliche Daten von Servern, die diese Bibliothek benutzten, ausgelesen. Das war möglich, weil die TLS-Antwort aus dem Wert „Payload-length" der Anfrage generiert wurde anstatt aus der tatsächlichen Länge der Payload und keine Grenzenprüfung durchgeführt wurde. Da für jeden Benutzer nur begrenzter Speicher allokiert wird, können durch eine überlange Payload darauffolgende Daten, welche meistens zu Datenobjekten anderer User gehören, in die TLS-Antwort gelangen. Im April 2014 wurde der Bug behoben, aber wie könnte eine weiterreichende Lösung zum Schutz vor Buffer-Overread bzw. Overflow aussehen?

MTD ist ein Konzept, welches dynamisch kontrollierbare Veränderungen eines Systems in einem Netzwerk ermöglicht. Wenn Systeme ihre nach außen sichtbaren Konfigurationen oft ändern, haben Angreifer wesentlich mehr Schwierigkeiten, in ein bestimmtes System einzudringen. Ein Cyberangriff funktioniert normalerweise so, dass die Opfersysteme zuerst ausgespäht werden und möglichst viel Information über diese gesammelt wird. Mit diesen Informationen kann dann ein gut auf die jeweiligen Systeme zugeschnittener Angriff gestartet werden. Ändern sich aber gewisse Konfigurationsparameter schon wieder bevor der wirkliche Angriff stattfindet, so greift dieser ins Leere. Das Angriffsfenster hat sich verschoben. MTD will sich den Vorteil der Zeit, die zwischen Ausspähung und Angriff vergeht, zunutze machen. Konfigurationsparameter können beispielsweise IP-Adressen, Namen, Netzwerke usw. sein.

2. MTD – Theorie
Zunächst erkläre ich theoretische Grundlagen, die Zhuang in [1] als Basis definiert.

2.1 Exploration surface
Das *Exploration surface* ist der Raum aller Ressourcen, die ein Angreifer durchsuchen kann, bzw. die an der „Oberfläche" nach außen sichtbar sind. Diese werden bei einer Cyberattacke genau erforscht und aus den gewonnenen Informationen bestimmt der Angreifer das *Attack surface* zu seinen Gunsten. MTD hat nun den Ansatz, das Exploration surface möglichst groß zu machen, sodass es einen großen Aufwand bedeutet, alle Ressourcen auszuspähen und so die Zeit der Angriffsvorbereitung deutlich zu erhöhen. Komponenten der Exploration surface können zum Beispiel IP-Adressen, aktive Software oder Ports sein.

2.2 Attack surface
Das *Attack surface*, von dem die Rede ist, wird auch Angriffsfenster genannt. Der derzeitige Ansatz von MTD ist es, das Attack surface so klein wie möglich zu machen und somit das System zu verhärten und sicherer zu machen. Alle verwundbaren Ressourcen, die einem Angreifer zugänglich sind, befinden sich in diesem Attack surface. Um auf eines der vorangegangen Beispiele einzugehen, könnten das Softwarebugs und –verwundbarkeiten sein.

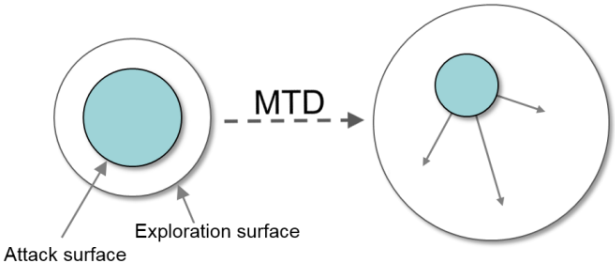


Abbildung 1: MTD Surfaces Konzept

Ziel ist es, den Angreifer möglichst stark zu „verwirren", indem das zu schützende System oft seine Erscheinung im Netz ändert und Angreifer das Ziel am besten aus den Augen verlieren. Das Attack surface wird „verschoben", was bedeutet, dass es einen

doi: 10.2313/NET-2015-03-1_12

Konfigurationswechsel durchmacht. Diese passieren durch Movement oder Transformation (=Adaptionen). Abbildung 1 zeigt links ein herkömmliches System und rechts ein MTD-System mit Bewegung bzw. Adaption des Attack surface innerhalb des Exploration surface.

2.3 Adaptionen und das konfigurierbare System

Die Verschiebung des Attack surface nennen wir Adaption. Diese bestehen entweder aus Bewegung, also Modifizierung einer bestimmten Konfigurationseinheit oder aus Transformation, worunter wir Veränderung der Anzahl der Konfigurations-einheiten verstehen. Adaption bedeutet, den Konfigurations-zustand eines MTD Systems in einen neuen validen Zustand mit veränderten Konfigurationsparametern zu transformieren.

Ein Konfigurationszustand beschreibt spezifische Zuordnungen von konkreten Werten an Konfigurationsparameter. Ein Beispiel wäre Arbeitsspeicher = 4 GB, HDD Größe = 500GB, IP Adresse 192.168.0.54 usw.

Konfigurierbare Systeme bestehen aus einer Menge von Konfigurationszuständen, in denen sich das System befinden kann, einer Menge an Aktionen, die es ausführen kann und einer Übergangsfunktion. Die Übergangsfunktion beschreibt die erlaubten Änderungen der Zustände.

In einem konfigurierbaren System gelten System Policies. Diese sind ein Regelwerk, das zur Prüfung von neuen Zuständen und Adaptionen auf Validität benutzt wird. Valide bedeutet hierbei, dass das System alle spezifizierten Ziele nach der Adaption noch erreichen kann. Diese unterteilt man in sicherheitstechnische und operationale Ziele. Die Sicherheitsziele bestimmen die kritischen Teile des Systems, die beschützt werden müssen, während die operationalen Ziele alle anderen Zwecke, die eine Plattform ausführen soll darstellt.

Ein MTD System besteht laut formaler Definition aus einem konfigurierbaren System, einer Menge an Policies und einer Menge an Zielen. Um einen möglichst hohen Grad an Unvorhersehbarkeit des zu beschützenden Systems zu erzeugen, soll hohe Diversifikation eingesetzt werden. Dies erreicht man durch einen sehr großen Raum an verschiedenen Konfigurationszuständen und durch eine Technik, die die Größe des Konfigurationsraums vergrößern soll. Letztere wird auch künstliche Diversifikation genannt. Als Faustregel gilt also: Je größer der Raum möglichst voneinander verschiedenen Zuständen ist, desto besser ist der Schutz vor Erraten.

2.4 Konfigurationswechsel

Zunächst werden ein neuer Konfigurationszustand aus dem Zustandsraum und eine dazu passende Adaption gewählt. Die Wahl derer kann innerhalb fester oder zufälliger Zeitintervalle erfolgen, wobei zufällige Zeitpunkte das System weniger vorhersehbar machen als feste. Man kann auch Umgebungsinformationen wie Intrusion Detection Alarme als Trigger für den Konfigurationswechsel hinzuziehen.

Sowohl der neue Zustand als auch die gewählte Adaption wird dann anhand der System Policies auf Validität geprüft. Schlägt diese Prüfung fehl, so wird eine neue Adaption bzw. ein neuer Zustand gewählt und die Prüfung wird erneut durchgeführt. Ansonsten wird sie ausgeführt. Abbildung 2 illustriert diesen Ablauf.

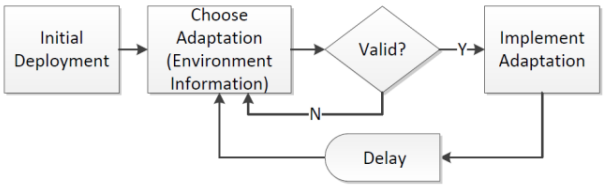


Abbildung 2: General MTD Process [1]

2.5 Randomisierung

Ziel der Randomisierung ist es, vollen Gebrauch der verfügbaren Konfigurationszustände und –parameter zu machen. Die Wahrscheinlichkeit, welcher Zustand als Nächstes gewählt wird, soll entweder gleichverteilt sein oder mit Hinzunahme von Alarmsystemen und zusätzlichen Informationen über potenzielle Angriffe und Verwundbarkeiten ein intelligenteres MTD System formen.

2.6 Probleme

2.6.1 MTD Problem

Eines der Kernprobleme ist es, welcher (valide) Konfigurations-zustand als Nächstes gewählt werden soll. Wie beim Punkt Randomisierung schon erwähnt, kann dies per zufälliger Auswahl erfolgen oder kombiniert mit intelligenteren Methoden wie Angriffserkennungssystemen (z.B. Intrusion Detection Systems) oder kostenbasierten Strategien.

2.6.2 Adaptionsauswahlproblem

Um zum nächsten Zustand zu kommen muss eine passende Adaption, welche aus einer Sequenz an Aktionen besteht gewählt werden. Der Streitpunkt besteht darin, dass es möglicherweise mehrere dieser Aktionssequenzen gibt und man die optimale Lösung finden muss. Hier sollen auch Zeit und Kosten berücksichtigt werden.

2.6.3 Timingproblem

Das richtige Timing, wann eine Adaption durchgeführt werden soll, ist ein besonders wichtiger Faktor für den Erfolg der MTD. Hierbei werden das Intervall zwischen den Adaptionen, die Zeit, die eine Adaption für einen Zustandswechsel benötigt, und die Zeit, die ein Angriff geschätzt benötigt, betrachtet. Die Entscheidung für den Zeitpunkt einer Adaption soll basierend auf dem Tradeoff zwischen Operation und Sicherheit getroffen werden. Je öfter eine Rekonfiguration stattfindet, desto höher sind die Performanceansprüche. Trotzdem sollte dies oft genug passieren, denn nur wenn sich die Konfigurationsparameter genau zwischen der Ausspähung durch den Angreifer und seinem Angriff ändern, kann das MTD System diesem entgehen.

doi: 10.2313/NET-2015-03-1_12

3. Weitere Konzepte

MTD ist, wie schon erwähnt, eine Technik für die Cybersicherheit und gilt als Beispiel für die Theorien der Systemagilität und des Effective Movement. Des Weiteren wird sie in diesem Abschnitt anhand eines spieltheoretischen Entwurfs modelliert.

3.1 Cybersicherheit und Diversität

Wir betrachten hier das Modell eines asynchronen Netzwerks mit n Knoten, welches z.B. Datenbankserver, Clienten und Router haben kann.

Nach Cybenko [2] bezieht sich MTD wie folgt auf die Cybersicherheit.

3.1.1 Kernziele der Cybersicherheit

Die Cybersicherheit lässt sich generell in drei Kernziele einteilen. Zuerst sollte ein Netzwerk Verfügbarkeit garantieren. Das bedeutet formal, dass mindestens einer von allen Knoten bzw. Systemen im Netzwerk nicht gefährdet ist. Des Weiteren sollte Integrität herrschen, was voraussetzt, dass die Mehrzahl, also höchstens die halbe Anzahl der Knoten, nicht gefährdet (oder bereits angegriffen) ist. Am besten jedoch sollte Vertraulichkeit erreicht werden. Das heißt, keiner der Knoten ist gefährdet. Gefährdet soll hier Zugang zu kritischen Informationen bedeuten.

Es gibt somit einen Tradeoff zwischen Vertraulichkeit und Verfügbarkeit. Im schlechtesten Fall trifft Verfügbarkeit, und im besten Fall Integrität und Vertraulichkeit zu.

3.1.2 Abhängigkeit

Abhängigkeit eines Systems bzw. Knotens von einem anderen resultiert, wenn eine Gefährdung des einen Knotens die Gefährdung des anderen zur Folge hat. Eine Zufallsvariable („time-to-compromise Variable“) misst die Zeit, die ein System von der Ausspähung des Exploration surface bis zum Angriff hat. Existiert Abhängigkeit zwischen zwei oder mehreren Knoten, so wird die Zufallsvariable für den Angriff auf den abhängigen Knoten maßgeblich von dem von ihm abhängigen Knoten beeinflusst. So werden nun zwei ähnliche bis gleiche Systeme in einem Netz mit hoher Wahrscheinlichkeit schneller mit nur kurzem Zeitabstand nacheinander angegriffen, wenn einer von beiden erfolgreich angegriffen wurde.

3.1.3 Diversität

Befinden sich im Netz nun nur äquivalente bzw. homogene Knoten, so spricht man von einer Monokultur. Angreifer können dort deutlich schneller weitere angreifen, wenn nur einer zum Opfer gefallen ist. Da hier aufgrund der Ähnlichkeit der Systeme zueinander die Exploration surfaces fast genau die gleichen sind, braucht der Angreifer keinen bis nur noch wenig Aufwand betreiben, um weitere Knoten auszuspähen, nachdem er einen einzigen erfolgreich erforscht hat. Dies ist z.B. eines der Konzepte von Botnets und Code Reuse. Monokulturen sind einfach zu skalieren und zu handhaben, büßen dafür aber Sicherheit ein. Die Zufallsvariablen für Angriffe der Knoten sind normalerweise nicht voneinander unabhängig.

Heterogene, also Knoten die verschieden voneinander sind, müssen alle einzeln mit erhöhtem Aufwand vom Angreifer ausgeforscht werden. Sie sind kostspieliger und schwieriger zu warten als Monokultursysteme, bieten aber erhöhte Sicherheit. Natürliche Diversität wird erreicht, wenn man ein heterogenes Netz hat, bei dem keine der Zufallsvariablen voneinander abhängig sind.

3.2 Systemagilität

3.2.1 Herausforderungen an die Systemagilität

Das ganze Konzept der MTD sollte so gut wie möglich vor dem Angreifer verheimlicht werden, also transparent sein. Das konkrete Verfahren sollte möglichst unvorhersehbar sein. Zur Prävention von Erraten durch den Gegner muss der Konfigurationsraum genügend groß gewählt werden.

Es ist außerdem eine Herausforderung, die Sicherheit aller Schichten des OSI-Modells aufrecht zu erhalten [3]. Die Konsistenz der Abhängigkeiten durch alle Schichten muss gewährleistet werden. Ein Beispiel wäre Multipath-Routing in einem Netzwerk. Bewegt man z.B. einen Dienst vom einen Knoten zum Nächsten, was natürlich von außen nicht sichtbar geschehen soll, muss man beachten, dass Angreifer TCP-Header mit Informationen über den neuen (und alten) Aufenthaltsort des Dienstes abfangen können.

Außerdem muss auch das Kostenmanagement berücksichtigt werden. Um Agilitätstechniken anzuwenden, müssen entweder zusätzliche Assets zu einem System installiert werden oder Performance im vorhandenen System eingebüßt werden.

3.2.2 Agilitätsmechanismen

Heutzutage sind noch die meisten Agilitätsmanöver reaktiv, was bedeutet, dass eine Verteidigungstechnik greift, nachdem ein Angriff festgestellt wurde. MTD soll proaktiv wirken, also präventiv Angriffe verhindern. Beispiele für Agilitäts-mechanismen sind Wrapperklassen, welche den Input für sicherheitsempfindliche Interfaces (z.B. System calls) umschreiben können und Verteilerdienste, die andere Dienste steuern (also allokieren, platzieren oder ausschalten) können. Weitere sind z.B. die Möglichkeit, Systemeigenschaften und –interfaces zur Laufzeit zu ändern und Täuschungstechniken, wo der Angreifer zur Erhöhung seines Aufwands gezwungen wird, indem man zuerst sein Verhalten beobachtet und ihn dann mit falschen Informationen füttert.

3.3 Effective Movement

MTD hat das Ziel, Systeme weniger deterministisch und statisch zu gestalten. Sind sie untereinander außerdem weniger homogen, wird der Aufwand, ein System erfolgreich anzugreifen, deutlich gehoben. Dies ist ein Beispiel für die Theorie des Effective Movement, deren Hauptleitfragen sich nach [4] folgendermaßen ergeben.

3.3.1 Umfang

Zunächst stellt sich die Frage, welche Elemente von der Technik abgedeckt werden sollen („Coverage“). Optimaler Weise sollte das gesamte Attack surface davon umschlossen werden und von der MTD Technik bewegt werden. Die Herausforderung, alle verwundbaren bzw. schützenswerten (also im Attack surface befindlichen) Komponenten im System zu finden, spielt dabei eine grundlegende Rolle.

doi: 10.2313/NET-2015-03-1_12

3.3.2 *Unvorhersehbarkeit*

Der Raum für die Bewegung muss hinreichend groß sein, um erfolgreiches Raten möglichst auszuschließen. Dieser wird durch die Kardinalität des Konfigurationszustandsraums bestimmt. Hierbei gilt: Je höher die Entropie einer Komponente, desto höher ist die Güte der Unvorhersehbarkeit. Die Entropie eines Knotens sinkt mit steigender Abhängigkeit der Knoten voneinander – je mehr Referenzen sie also untereinander haben.

3.3.3 *Rechtzeitigkeit*

Das Zeitmanagement der MTD spielt ebenso eine entscheidende Rolle für die Qualität der Cyberabwehr („Timeliness“). Die Konfigurationswechsel sollten am besten genau zwischen der Ausspähung des Systems durch den Angreifer und dem eigentlichen Angriff erfolgen. Es ist nützlich, viele Informationen über den Angreifer zu sammeln und daraus - beispielsweise durch einen Machine-Learning-Algorithmus - ein „Attacker model“ zu generieren, anhand dessen man seine Verteidigungsstrategie optimal anpasst. Hier stellt sich das Problem, dass diese Ausspähung von der Seite des Verteidigers aus schwierig zu erkennen ist. Fehlen solche Informationen, kann man den Zeitplan nicht an die Bedrohung anpassen und es liegt nahe, dass man die Zustandsbewegung möglichst oft vollzieht und dabei blind hofft, einem Angriff zuvorzukommen.

Weitere wichtige Kriterien für die Effektivität sind unter anderem der resultierende Overhead in Performanz, Speicher oder Netzwerk, die Kosten für z.B. Entwicklung, Einsatz, und Wartung, die Expertise, die auf menschlicher Ebene notwendig ist um mit dieser neuen Technik umzugehen und die Kompatibilität mit den bereits vorhandenen Komponenten.

3.4 MTD als spieltheoretischer Ansatz

3.4.1 *Modell*

Die Interaktion zwischen Angreifer und Verteidiger bei MTD kann man gut als Leader-Follower-Spiel mit zwei Spielern modellieren [5]. Abbildung 3 zeigt den Aufbau des Spiels. Das „Spielfeld“ sind die Plattformen, welche abwechselnd von Spielern „besetzt“ werden. Diese simulieren bei der MTD die Konfigurationszustände und deren Parameter haben jeweils unterschiedliche Werte. Es kann immer nur genau eine Plattform gespielt werden, was übertragen auf die Realtität bedeutet, dass ein Konfigurationszustand gerade auf einem System aktiv ist. Ein Knoten ist verwundbar/angreifbar, wenn er mindestens eine ausbeutbare Eigenschaft hat (rote Kreise im Bild). Besetzt der Verteidiger gerade einen verwundbaren Knoten und der Angreifer wählt diesen für seinen nächsten Spielzug, so wird er angegriffen.

Der Verteidiger ist der Leader im Spiel, also macht seinen Spielzug als Erster. Er weiß nicht, welche Knoten im Netz angreifbar sind und hat keine Möglichkeit herauszufinden, welche Knoten angegriffen werden - er sieht die Spielzüge des Angreifers also nicht. Er kann einmal befallene Knoten auch nicht wieder zurückgewinnen. Sein Ziel ist es, persistenter Bedrohung über die Dauer des Spiels vorzubeugen.

Der Angreifer ist der Follower und besitzt komplette Information darüber, welche Knoten angreifbar/verwundbar sind. Er hat eine begrenzte Menge an Exploits als Angriffsmittel, die jeweils für bestimmte Plattformen greifen und kann ausschließlich den momentan aktiven Knoten angreifen.

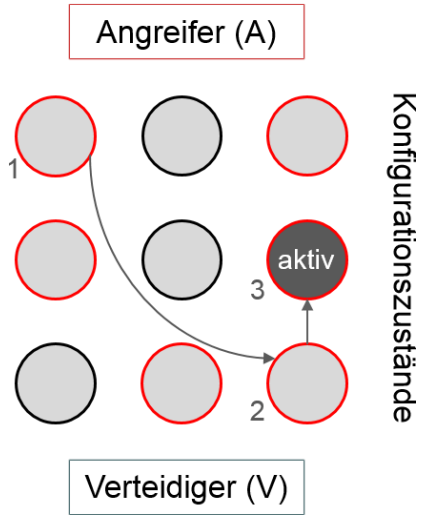


Abbildung 3: Spielfeld mit 3 Zügen nacheinander. Rote Felder: verwundbare Zustände, schwarze unverwundbare.

Spielt der Verteidiger über eine bestimmte Zeitdauer immer verwundbare Knoten, für die der Angreifer einen Exploit hat (und den auch anwendet), resultiert eine persistente Bedrohung und der Angreifer kann gewinnen. Die zeitliche Dauer kann z.B. in Minuten oder in der Anzahl der Spielzüge gemessen werden.

Das Spiel kann nun im Modus „Crash the Satellite“ oder im „Exfiltration Over a Difficult Channel“ gestartet werden. Hat der Angreifer bei „Crash the Satellite“ für eine bestimmte Dauer Kontrolle über das System (= persistente Bedrohung), so hat er gewonnen. Bei „Exfiltration Over a Difficult Channel“ muss der Angreifer über eine bestimmte Zeitdauer ein verwundbares System besetzen, damit sein Angriff erfolgreich ist. Ist das geschehen, so steigt seine Payoff-Funktion mit der Zeit.

3.4.2 *Wahl der Strategien*

Der Verteidiger geht intuitiv davon aus, dass der letzte Knoten, den er gespielt hat, angegriffen wurde und verwundbar war. Je ähnlicher ein System einem anderen ist, desto wahrscheinlicher teilen sie sich auch eine oder mehrere Verwundbarkeiten. Deshalb wird der Verteidiger als Nächstes einen möglichst unterschiedlichen spielen.

Der Angreifer kann auf zwei Arten operieren: im statischen Modus hat der Angreifer bereits am Anfang alle Fähigkeiten und gewinnt nichts durch das Beobachten des Systems. Hier ist das einzige Ziel des Verteidigers, die Wahrscheinlichkeit, dass der Angreifer eine verwundbare Plattform trifft, zu minimieren.

Im adaptiven Modus kann er über Zeit Gegenmaßnahmen entwickeln, je mehr Informationen er durch das Beobachten des Netzes, also die Züge des Verteidigers gewonnen hat. Als Beispiel könnte man hier einen Spam-Filter nennen, wo ein Angreifer nach jedem Senden einer Mail etwas über die Eigenschaften des Filters lernt und Antwortmails generieren kann. Der Verteidiger muss nun sowohl Angriffe abwehren als auch die Wahrscheinlichkeit, dass der aktuelle Knoten getroffen wird, minimieren.

doi: 10.2313/NET-2015-03-1_12

3.4.3 *Ergebnisse*

Die statistisch optimale Lösung nach der Auswertung von verschiedenen Gefahrenmodellen resultiert, wenn die Plattformwahl präferenziell geschieht - wohingegen rein zufällig gewählte Knoten nicht die klügsten Entscheidungen sind. Bezogen auf MTD heißt das, dass Verteidigung am besten klappt, wenn die Konfigurationen immer so verschieden zur vorherigen gewählt werden, wie es geht. Allerdings bleibt zu berücksichtigen, dass die optimalen Moving Target Strategien immer vom Modell abhängen, also möglichst individuell auf das Vorgehen des Angreifers eingegangen werden muss.

4. Beispiele

4.1 Address Space Layout Randomization

Address Space Layout Randomization (ASLR) ist laut [4] das derzeit verbreitetste Beispiel für MTD und wird z.B. eingesetzt, um der Bedrohung von Code Reuse und Code Injection Angriffen entgegenzuwirken. Wenn Angreifer Speicherorte von nützlichen Datenobjekten und Code einer Software herausfinden, können sie mit darauf zugeschnittenen Exploits jedes System schädigen, das diese Software benutzt, da die Speicheradressen überall die gleichen festen Offsets haben.

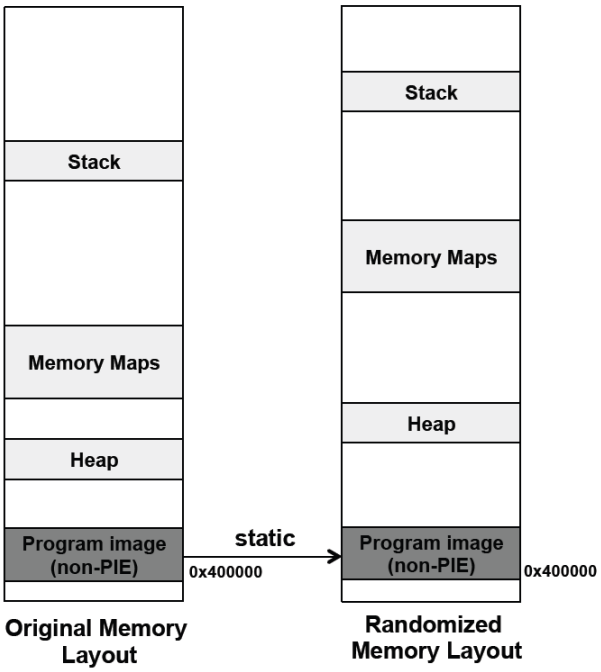


Abbildung 4: Speicherlayout ohne positionsunabhängigen Executables (PIE) [4]

Optimaler Weise sollen also auch die Program executables eine dynamische Speicheradresse bekommen. So wären alle Programmteile von der MTD-Technik abgedeckt und zusammen mit einer klugen Wahl der neuen virtuellen Speicheradressen wären keine davon leicht vorhersehbar.

ASLR löst die Programmteile von ihren statischen Adressen im Speicher, sodass sie dynamisch geändert werden bzw. an zufälligen Speicherorten stehen und das Speicherlayout also auf jedem System anders aussieht. Die meisten ASLR-Implementationen randomisieren den Stack, Memory Maps und den Heap (siehe Abbildung 4), aber das Program image ist oft noch positionsabhängig (PIE), da positionsunabhängige (non-PIE) Programme gesondert kompiliert werden müssen.

Ein Problem, das sich bei dieser Technik leider seit einiger Zeit ergibt, ist das *Spraying*, wo Schadcode ähnlich einer Spraydose großflächig in den Speicher gesprüht und dupliziert wird. So soll die Wahrscheinlichkeit steigen, dass trotzdem irgendwann ein Bibliotheksaufruf durch den eingeführten Code geschieht.

4.2 Spatio-temporal Address Mutation

Hierunter versteht man das dynamische Mapping von Hosts an IP-Adressen [6]. Als Konfigurationsparameter werden in diesem MTD-Konzept also IPs verwendet. Die Basis besteht darin, dass jeder Host, der mit einem anderen interagieren will, eine kurzlebige (engl. ephemeral) eIP zugewiesen bekommt, mit der er dann ausschließlich mit dem angegebenen Partner kommunizieren darf.

Ein sogenannter Controller übernimmt die Mappings der (realen) rIPs auf eIPs pro Host, bestimmt die Mutationsstrategie, autorisiert Zugriffe mit rIPs (welche nur von Administratoren ausgeführt werden dürfen) und schickt die Mapping-Tabelle an Gateways. Diese sind für die Transformation der IPs selbst verantwortlich – übernehmen also die „Arbeit“. Sie prüfen die IPs jedes Pakets, das passieren will, anhand der Mapping-Tabelle auf Legalität und transformieren die rIPs anschließend in eIPs und umgekehrt, oder verwehren ungültigen Paketen den Weg.

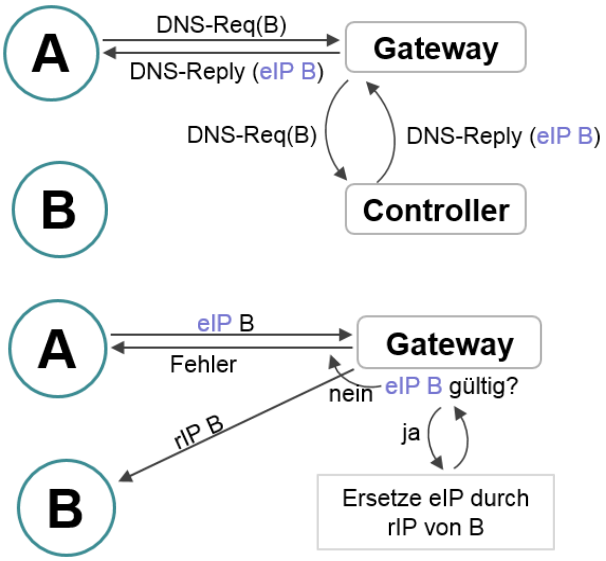


Abbildung 5: Kontaktaufnahme (oben) und Kommunikation (unten) bei STAM

Zur Kontaktaufnahme muss ein Host zunächst ein DNS-Request für den Empfänger starten. Der Controller erstellt nach dessen Eingang ein neues Mapping der Empfänger-rIP an eine eIP,

doi: 10.2313/NET-2015-03-1_12

welche mit der DNS-Reply zurück an den Anfragesteller-Host geht. Dieser verwendet daraufhin die neue eIP für jedes Paket, das er an diesen Partner schickt. Nachdem das Gateway die Gültigkeit der Pakete validiert und die Transformation der eIP zur rIP des Empfängers vorgenommen hat, schickt es sie weiter. Abbildung 5 visualisiert diese Vorgänge vereinfacht.
Die eIPs sind kurzlebig, weil sie vom Controller einen time-to-live (TTL) Wert als Verfallszähler bekommen, was die temporale Komponente des Konzepts darstellt. Die Wahl des TTL-Werts ist gemäß des Poisson-Erneuerungsprozesses. Die spatiale (räumliche) Komponente ist die Entscheidung, welche eIP als nächstes verwendet werden soll. Hierzu kann die Wahl entweder uniform oder täuschend sein. Bei der uniformen Wahl nimmt der Controller irgendeine aus den verbleibenden eIPs, wobei jede IP die gleiche Wahrscheinlichkeit hat. Bei der täuschenden Mutation nimmt der Controller die eIP, für welche er einen Angriff am unwahrscheinlichsten hält.

4.3 Field Programmable Gate Arrays

Field Programmable Gate Arrays (FPGAs) sind rekonfigurierbare Hardwarekomponenten bzw. Rechenbausteine, die dabei helfen, kritische Programmabschnitte auf Hardwareebene zu partitionieren (*Hardware Partitioning)* und somit zu beschützen. FPGAs sind weitaus primitiver gebaut und daher auch billiger als normale CPUs.
Dadurch, dass die FPGAs individuell konfigurierbar sind, kann der Softwaredesigner selbst entscheiden, welche Teile des Programms auf den FPGAs laufen sollen. Der wichtigste Vorteil hiervon ist die Verbesserung der Performanz des Systems. Implementiert man ein Sicherheitsprogramm z.B. für ein FPGA, so wird zu dessen Ausführung keine CPU-Zeit in Anspruch genommen, wodurch man also Parallelität hergestellt hat.
Besonders praktisch bezüglich Sicherheitsrisiken sind sie, wenn man im Adressraum von Programmen springen muss. Man kann das Programm in zwei Bereiche aufteilen, sodass die gefährdeten Teile des Programms auf einem FPGA und die sicheren Teile auf einer (oder mehreren) herkömmlichen CPU implementiert werden. Moving Target Defense ist hier also die Auslagerung bzw. Bewegung von Programmteilen innerhalb der Rechnerarchitektur [7].
Ein spezieller Ansatz für den Einsatz von FPGAs für Datensicherheit ist *CODESSEAL*. Hier befinden sich Programm und Daten zur Laufzeit verschlüsselt im Speicher. Zwischen dem Cache und dem Speicher befindet sich ein FPGA zur Ver- und Entschlüsselung der Daten, die von der CPU verarbeitet werden. Auf diese Weise befinden sich nur verschlüsselte Daten zwischen externer Hardware und Prozessor, also können Adressen und Datenleitungen nicht gesnifft werden.

5. Zusammenfassung

Man kann also folgern, dass MTD eine zukunftsorientierte Technik ist. Es soll nicht mehr still darauf gewartet werden, bis man von einer Cyberattacke angegriffen ist um sich dann erst zu versuchen zu verteidigen, sondern sich schon im Voraus durch geschickte Veränderungen der im Netz bzw. von Angreifern sichtbaren Systemressourcen so unkenntlich wie möglich zu machen.
Je höher der Grad der Diversifikation und der Randomisierung, desto geringer ist die Wahrscheinlichkeit, dass der Angreifer den nächsten Konfigurationszustand voraussagen kann. Dies kann durch zufällige Wahl oder besser noch kombiniert mit Angriffserkennungssystemen erreicht werden.

Als Schutz vor Buffer-Overflows wird heutzutage oft eine Technik eingesetzt, die einzelne Pages bzw. Datenobjekte im Speicher schützt und einen IDS-Alarm gibt, sobald geschützte Pages von User-Anfragen berührt werden [8]. Abbildung 6 zeigt ein solches Schema. Mit dieser Schutztechnik können nicht alle Pages gleichzeitig abgedeckt werden, da sich die Antwortzeit des Servers proportional zur Anzahl bewachter Pages verhält.
Hier wäre es nun eine MTD Idee, den Schutz der Pages dynamisch infolge von bereits erfolgten Angriffen anzupassen. So könnte man z.B. unter Einbezug von Machine Learning Algorithmen mehr Pages in der Nähe derer schützen, die schon einmal angetastet wurden.

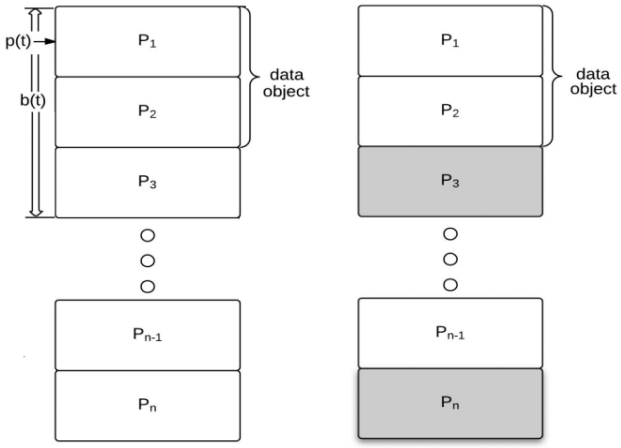


Abbildung 6: Speicherlayout ohne Page-Schutz (links) und mit Page-Schutz (rechts). b(t) = ausgelesene Pages durch das Payload-length-Feld; p(t) = erste zu lesende Page.

6. Referenzen

[1] R. Zhuang, S. A. DeLoach, X. Ou. Towards a Theory of Moving Target Defense, 2014.

[2] G. Cybenko, J. Hughes. No Free Lunch in Cyber Security, 2014.

[3] P. McDaniel, T. Jaeger, T. F. La Porta, N. Papernot. Security and Science of Agility, 2014

[4] T. Hobson, H. Okhravi, D. Bigelow. On the Challenges of Effective Movement, 2014.

[5] K. M. Carter, J. F. Riordan, H. Okhravi. A Game Theoretic Approach to Strategy Determination for Dynamic Platform Defenses, 2014.

[6] J. H. Jafarian, E. Al-Shaer, Q. Duan. Spatio-temporal Address Mutation for Proactive Cyber Agility against Sophisticated Attackers, 2014.

[7] T. R. Andel, L. N. Whitehurst, J. T. McDonald. Software Security and Randomization through Program Partitioning and Circuit Variation, 2014.

[8] M. Zhu, Z. Hu, P. Liu. Reinforcement Learning Algorithms for Adaptive Cyber Defense against Heartbleed, 2014

doi: 10.2313/NET-2015-03-1_12

Cryptocurrency Brings New Battles into the Currency Market

Yingjie Zhao
Betreuer: Heiko Niedermayer
Seminar Future Internet WS2014
Lehrstuhl Netzarchitekturen und Netzdienste
Fakultät für Informatik, Technische Universität München
Email: yingjie.zhao@in.tum.de

ABSTRACT

In this paper, we concern ourselves with cryptocurrency and how cryptocurrency affects the cryptocurrency market as well as the fiat currency market. The whole topic will be sepreted into two sections: competition among different currencies, as well as competition among exchanges[2]. We aim at figuring out the current circumstance of cryptocurrency which as a casual visitor in the market, and additionaly we will also look at the prospect of cryptocurrency and the currency market. Cryptocurrency with many new features has an uneasy development after entering into the financial market, although it is not yet powerful to compete with fiat currency, the effects of cryptocurrency and cryptocurrency exchange in financial market will still be full of meaning.

Keywords/Schlüsselworte

Cryptocurrency, Bitcoin, fiat currency, digital wallet, cryptocurrency exchanges

1. INTRODUCTION

Even though the first decentralized digital currency Bitcoin is created in 2009, it caught the interest of the mainstream media only in 2012[2]. Because of several special features cryptocurrencies are often compared to fiat currencies, hence, the first major part of this paper is to show the current developments within the cryptocurrency system, the competition within different cryptocurrencies and with fiat currencies as well. Such developments and competition may have an important impact not only on the future development of currency market but also on the success of technology innovation. In this part, we first have a look at the difference among the most important or successful cryptocurrencies and their differences. Apart from the competition among cryptocurrencies, the battle between cryptocurrency and fiat currency will also affect the financial market. Hereby the advantage and disadvantage of each currencies will be analyzed one by one, we expect that Bitcoin is the most competitive cryptocurrency, we also find that cryptocurrency is still in infancy so that it can not compete against or even replace fiat currency, and finally we figure out a conclusion. In the second major part of this paper, we move our focus to competition among cryptocurrency exchanges and competition between digital wallet and bank. The market is changing, as well as the top exchange in the cryptocurrency market, although OKCoin seems to be the best exchange today, it is still hard to tell whether OKCoin will keep this status for a long time or not. Additionally, there are some factors which effect the competition among exchanges. Except from cryptocurrency exchange, we introduce here a new idea — the digital wallet, i.e. a wallet of digital currencies or a bank of digital currencies, where people can take control of their coins like transfer their coins overall the internet. At the end of this part, we figure out the strength of the digital wallet in comparison with bank and its present situation in the currency market.

2. BRIEF BACKGROUND OF CRYPTOCURRENCY

2.1 Definition

In Wikipedia, cryptocurrency is defined as a medium of exchange using cryptography to secure the transactions and to control the creation of new units. Cryptocurrencies are a subset of alternative currencies or specifically of digital currencies[5].

2.2 Features

The first common feature in cryptocurrency is decentralized control, which means they have no central authority[2], they are distinct from a centralized electronic money system such as Paypal. Instead, they use cryptography to control transactions, increase the supply and prevent fraud[2]. Another common feature is that transactions are publicly recorded in a ledger. An example is bitcoin, where all transactions are recorded in the block chain[5].

2.3 Brief Historical Background

The first cryptocurrency was Bitcoin(BTC), it was created in 2009 by a pseudonymous developer named Satoshi Nakamoto. One of Satoshi Nakamoto's works is the most famous article named "Bitcoin: A Peer-to-Peer Electronic Cash System". With the publication of this article, Bitcoin as the first cryptocurrency was known more and more widely.

Apart from Bitcoin, there are also many other cryptocurrencies which named altcoin altogether. For example, Namecoin(NMC), the first altcoin was created in April 2011 which was due to form a decentralized DNS to make internet censorship more difficult[1]. And soon in October 2011, Litecoin(LTC) was released and became the first successful cryptocurrency to use scrypt as its hash function rather than SHA-256[1]. Thereafter more and more different kinds of

doi: 10.2313/NET-2015-03-1_13

cryptocurrency came into the market such like Peercoin(PPC), Feathercoin(FTC), Novacoin(NVC), Terracoin(TRC) and so on.

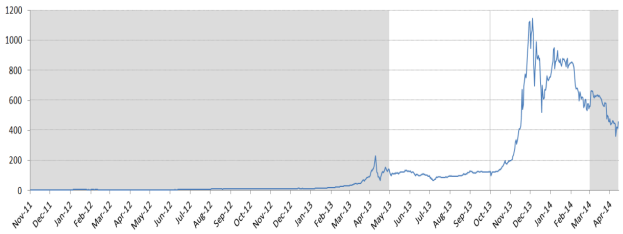


Figure 1.1 Bitcoin prices (in USD) over time. Shading highlights the first period (May–Sep 2013) and the second period (Oct 2013–Feb 2014) of our analysis. (Source: http://www.coindesk.com/price)

During the development of cryptocurrency there are two significant periods (Figure 1). The first period is from 2 May 2013 to 30 September 2013, and the second period is from 1 Oct 2013 to 28 February 2014. Since May 2013, Bitcoin's price to USD had a rise, and in Oct 2013, Bitcoin's price began to skyrocket especially after the shutdown of silkroad.com. Silkroad is a online balck market, where merchants are allowed to make deals using Bitcoin. After the shutdown of Silkroad, about 26 000 Bitcoins in worth of 3 200 000 Dollars are confiscated by government, and this let to reduction of the supply of Bitcoin, which causes the increase of Bitcoin's price. In Dec 2013 Bitcoin's price reached $1,240 at the peak. In February 2014 $350 million worth of Bitcoins were stolen from Mt.Gox which causes the loss of confidence in Bitcoin and thereafter Bitcoin's price began to fall continuously. Nowadays, Bitcoin's price to USD is around $350/BTC[3]. Additionaly, the price of Bitcoin is very volatile or unstable in the second period, some factors or events affect the price of Bitcoin in this period, for example the quick development of altcoins, or people have a new aspect on Bitcion and so on.

3. CRYPTOCURRENCY AS A NEW COMPETITOR IN THE CURRENCY MARKET

3.1 Cryptocurrencies

In this part, we first focus on major cryptocurrencies, the brief introduction of Bitcoin, the difference of major cryptocurrencies, their features and their competitions.

3.1.1 Bitcoin

As mentioned above, Bitcoin is the first and the most successful cryptocurrency, it gains the most popularity in the cryptocurrency market and media coverage has mostly focused on Bitcoin either. Bitcoin as a leader of cryptocurernncy has several features and advantages. Firstly, a bank account is not necessary. Secondly, it enables anonymous purchase. Thirdly, it wholly replaces state-baked currencies with a digital version which is tougher to forge. Bitcoin enable to cut across international boundaries, can be stored on personal hard drive instead of in a bank, and won't be easily manupulated by Federal Reserve [6].

Except these common features, Bitcoins are scarce as well, but their scarcity is algorithmic. New Bitcoins are added only by "mined", and there are also only a certain amount of Bitcoins — 21 million — can be mined, and that's why Bitcoin won't get the trouble with inflation theoretically, and people now are used to invest Bitcoins as an asset in the late period.

3.1.2 Other Major Cryptocurrencies and Difference

In fact, many of the altcoins like Litecoin and Peercoin were developed to fix the shortcomings of Bitcoin[2]. In the early period, these altcoins attract only a few users, but lately as people's view to digital coin has changed, there has been a surge in entry into the digital coin market, and since then people has begun to invest different sorts of digital currencies.

Except Bitcoin, the most successful cryptocurrency is Litecoin. It was firstly created on october 7th, 2011 and can be mined using Scrypt — another algorithm, and there are 84 million litecoins in total, that means Litecoin generates four times as many coins, and the transactions are added to the block chain four times faster than Bitcoin[2]. Such advantages make Litecoin the most competitive altcoin against Bitcoin.

Peercoin, another important altcoin. It relies on proof-of-stack in addition to proof-of-work to record transactions in the blockchain[2]. The most difference to Bitcoin or Litecoin is that Peercoin does not have a limit on the total number of coins generated[2]. So merchants do not have to worry that Peercoin will be all mined one day, but it may cause inflation which will affect the price of Peercoin as well as its store of value.

3.1.3 Competition among Cryptocurrencies

Market capitalisation values for different "coins" are quite skewed, total market capitalisation in digital currencies was approximately $8.1 billion till 26 February 2014. Bitcoin accounts for approximately 90% of total digital currency market capitalization. Litecoin takes the second place with approximately 5% and Peercoin accounts for only 1% of total market capitalization[2]. As the interest of cryptocurrency has grown quickly especially in 2013, more new cryptocurrencies entered into the market, so the competition between different kinds of cryptocurrencies became more intense. As the article "Competition in the Currency Market" postulates, there are two motivations for the introduction of new cryptocurrencies — fixing shortcoming of Bitcoin and capitalising on potential popularity [2]. The character of cryptocurrencies is not only pure as a currency but also as financial assets. Refers to the article "Competition in the Cryptocurrency Market", it postulates that cryptocurrencies play both roles nowadays in the market and with different effects.

The first effect is the reinforcement effect. This is the result of the one-sided network-effect, which means if something becomes more popular, then more people will believe in it, and such things like Bitcoin will win the "winner-take-all", this means it always gains more and more popularity. The second effect is the substitution effect. This is the result of speculative dynamics. This means, merchants in the cryptocurrency market regard cryptocurrencies as financial assets, they invest in other cryptocurrencies than Bitcoin because of their fear of volatility of Bitcoin. They do not want to put all their eggs in one basket.

doi: 10.2313/NET-2015-03-1_13

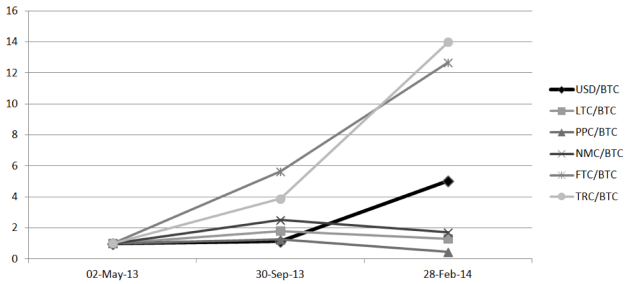


Figure 2. Changes in exchange rates of various currencies against BTC over the threshold dates in our data. Rate on 2 May 2013 is normalized to 1.

Through some data (Figure 2) we can see changes in prices which are good measures of changes in demand, and we can tell which effect is more dominant in each period.

In the first period (from 2 May 2013 to 30 September 2013), the value of all cryptocurrencies against Bitcoin increases relatively more slowly and gently in comparison with the second period. When Bitcoin becomes more valuable against USD, it also becomes more valuable against other cryptocurrencies. Additionally, Peercoin retains its value over the first period very stable, its value has hardly any significant changes against Bitcoin in the first period, and so do Litecoin and Namecoin, so we call them successful cryptocurrencies. In the second period, Feathercoin and Terracoin have significant loss in their value against Bitcoin, but at the same time, those successful coins in this period increase their value against Bitcoin, because since the second period merchants increasingly accept altcoins like Litecoin as an alternative to Bitcoin in order to reduce investment risk of Bitcoin, this means the demand of such altcoins raises more quickly than Bitcoin, so their price raises gainst Bitcoin as well. On the other side, such behavior of merchants also indicates that the financial function of cryptocurrencies becomes more prominent, so the substitution effect seems to be more dominant in this period.

Except from development trend of cryptocurrencies, the market capitalization is another proof of the competition among cryptocurrencies. According to the statistic of "24 Hour Volume Ranking"(on 10.12.2014)[9], we can see that Bitcoin has the most market capitalization, with 80.26%, Litecoin with 10.58% and Ripple with 3.80%. They are three major cryptocurrencies in trades with more than 90% market capitalization.

1. Bitcoin (80.26 %)

#	Source	Pair	Volume (24h)	Price	Volume (%)
1	Bitfinex	BTC/USD	$ 7,130,960	$ 349.56	32.52 %
2	OkCoin Intl.	BTC/USD	$ 5,391,550	$ 348.34	24.59 %
3	Bitstamp	BTC/USD	$ 4,336,740	$ 348.60	19.78 %
4	BTC-E	BTC/USD	$ 1,906,900	$ 344.46	8.70 %
5	LakeBTC	BTC/USD	$ 1,590,710	$ 351.46	7.25 %
6	Kraken	BTC/EUR	$ 705,984	$ 353.64	3.22 %
7	itBit	BTC/USD	$ 329,682	$ 350.50	1.50 %
8	BTC38	BTC/CNY	$ 142,871	$ 347.73	0.65 %
9	HitBTC	BTC/USD	$ 118,531	$ 348.95	0.54 %
10	BTC-E	BTC/RUR	$ 97,220	$ 346.24	0.44 %

Figure 3. "24 Hour Volume Ranking" – Bitcoin/Litecoin (Accessed on 10.12.2014) (Source: http://coinmarketcap.com/currencies/volume/24-hour/).

2. Litecoin (10.58 %)

#	Source	Pair	Volume (24h)	Price	Volume (%)
1	OkCoin Intl.	LTC/USD	$ 2,047,790	$ 3.52	70.83 %
2	BTC-E	LTC/BTC	$ 287,866	$ 3.51	9.96 %
3	Bitfinex	LTC/BTC	$ 163,968	$ 3.52	5.67 %
4	BTC-E	LTC/USD	$ 126,989	$ 3.46	4.39 %
5	BTC100	LTC/CNY	$ 67,908	$ 3.51	2.35 %
6	Bitfinex	LTC/USD	$ 64,132	$ 3.53	2.22 %
7	Virtex	LTC/USD	$ 28,538	$ 3.51	0.99 %
8	Cryptsy	LTC/BTC	$ 21,370	$ 3.49	0.74 %
9	HitBTC	LTC/BTC	$ 14,369	$ 3.53	0.50 %
10	BTC38	LTC/CNY	$ 14,295	$ 3.50	0.49 %
		View More			
Total/Avg			**$ 2,891,060**	**$ 3.52**	

Figure 4. "24 Hour Volume Ranking" – Bitcoin/Litecoin (Accessed on 10.12.2014) (Source: http://coinmarketcap.com/currencies/volume/24-hour/).

3. Ripple (3.80 %)

#	Source	Pair	Volume (24h)	Price	Volume (%)
1	Ripple Charts	XRP/BTC	$ 248,985	$ 0.015510	23.99 %
2	BTC38	XRP/CNY	$ 227,961	$ 0.015725	21.97 %
3	Ripple Charts	XRP/JPY	$ 216,041	$ 0.015953	20.82 %
4	Ripple Charts	XRP/USD	$ 205,416	$ 0.015489	19.79 %
5	Ripple Charts	XRP/CNY	$ 67,294	$ 0.015498	6.48 %
6	Cryptsy	XRP/BTC	$ 24,091	$ 0.017117	2.32 %
7	Poloniex	XRP/BTC	$ 17,471	$ 0.016049	1.68 %
8	Kraken	XRP/BTC	$ 15,739	$ 0.016228	1.52 %
9	AllCoin	XRP/BTC	$ 14,695	$ 0.016049	1.42 %
10	Cryptsy	XRP/USD	$ 84	$ 0.016002	0.01 %
		View More			
Total/Avg			**$ 1,037,778**	**$ 0.015710**	

Figure 5. "24 Hour Volume Ranking" – Ripple (Accessed on 10.12.2014) (Source: http://coinmarketcap.com/currencies/volume/24-hour/)

Although merchants accept more altcoins than ever before, Bitcoin is still no doubt their first choice, its scarcity and store of value attracts a lot of merchants. The reinforcement effect makes Bitcoin retain the largest market capitalization, and the substitution effect enables a quick development of altcoins. Under both effects the competition in the cryptocurrency market seems to be more unpredictable, maybe someday they will find a balance.

Generally, Bitcoin has an absolute superiority over other cryptocurrencies not only in price but also in market capitalization, it keeps its top position all the time, and it will probably keep this position for a very long time in the future. Even though after Mt.Gox incident, people prefer to treat Bitcoins as financial assets rather than as currencies.

3.2 Cryptocurrencies and Fiat Currencies

The appearance of cryptocurrencies, especially Bitcoin, has more or less impact on fiat currency market. On the one hand, the impact may be positive, on the other hand, it may be negative, accurately, the entry of cryptocurrencies into the market is a double-edged sword.

When Bitcoin's value suddenly skyrocketed in 2013, people pressed forward to purchase Bitcoins. Now we wonder, if cryptocurrencies will replace fiat currencies someday. Obviously, there is no certain answer, but at least the relationship between cryptocurrencies and fiat currencies should not be ignored. At the same time, as cryptocurrencies become more and more familiar to people, the relationship between cryptocurrencies and fiat currencies will get more and more closer. At this moment, cryptocurrency as a new member enter into financial market, it seems to be a strong competitor against

fiat currencies. In my opinion, although there is competition between cryptocurrencies and fiat currencies, in general, they could also be able to coexist as mutual optimisations of one another. In another word, cryptocurrencies do not have to replace fiat currencies in order to be successful, I think the "win-win" should be the best common goal to achieve.

3.2.1 Cryptocurrency is a New Competitor

As a new competitor to fiat currency, cryptocurrency has several advantages which attack a lot of merchants at the beginning.

At first, merchants can transfer their coins directly to another person without bank or clearinghouse. This means merchants can minimize their fees in transaction. Sometimes they may pay fees with transaction to receive priority processing, but this payment is still very low. This advantage also benefits those people who do not have a bank account but want to transfer their money. The second advantage is the free payment, merchants can send and receive any amount of money anywhere in the world. No borders, no limits, efficient, merchants can be in full control of their coins. This is rather attractive for people who usually transfer money abroad. Thirdly, there are fewer risks for merchants, because cryptocurrency transactions are secure, irreversible and do not contain any users' sensitive information, and this protects merchants from losses by fraud. Additionally, merchants can protect their money with backup and encryption, and another point, because cryptocurrency is decentralized, the price of each cryptocurrency will not be controlled or influenced by government or Federal Reserve, and cryptocurrencies do not have inflation because of its certain supply amount.

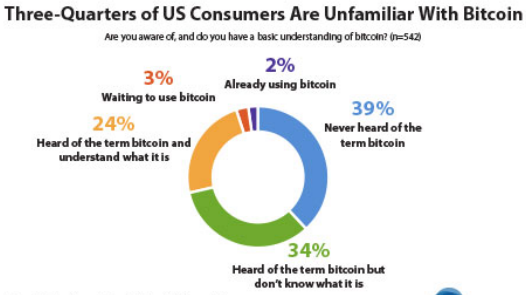


Figure 6. [4]

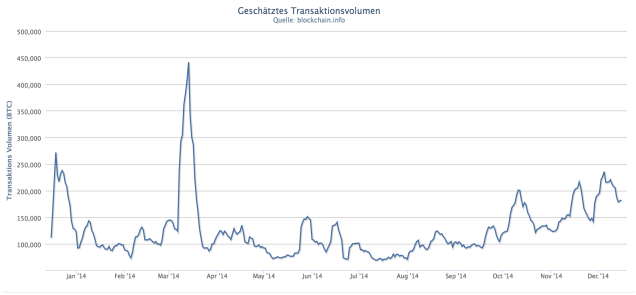


Figure 7. Estimated Transaction Volume. (Source: https://blockchain.info/de/charts/estimated-transaction-volume)

Except from these advantages, cryptocurrencies have disadvantages as well. There are still a lot of people unaware of cryptocurrencies, although more businesses accept them because they want the advantages of doing so. Cryptocurrencies are not manipulated by government, but this does not mean that the price of cryptocurrencies is very stable. In contrast, cryptocurrencies are very volatile. The total amount of cryptocurrencies in circulation and the number of businesses using cryptocurrencies are still very small. Therefore, even small events, trades or business activities can significantly affect the price[14]. If the adoption increases, the volatility will reduce. Furthermore, cryptocurrencies are still in infancy, they are currently in an ongoing development with many incomplete features. The Figure 6 also shows a very los acceptance of Bitcoin, which is actually a significant waekness of Bitcoin. There are too many things like security and services have to be improved, and the cryptocurrency business is very new. In another word, it is quite risky to do business with cryptocurrency, with the risks like policy risk, legal risk, speculative risk, money laundering and so on. The shutdown of Silkroad is a sign for warn and such an event brought a very negative effect on cryptocurrencies.

During the following long-term development, there are many unpredictable risks, it may be security issue like theft or fraud, it may be regulatory from government and some policies against cryptocurrencies.

At this moment, it is rather hard to tell the prospect of cryptocurrencies or to distinguish whether cryptocurrencies will be better than fiat currencies in the future, but there is one thing for sure, that cryptocurrencies have slowly become a ~~the~~ competitor to fiat currencies.

3.2.2 Cryptocurrencies vs. Fiat Currencies

Cryptocurrencies are now a competitor to fiat currencies no matter if they are powerful enough. Since the existence of cryptocurrencies, they affect more or less on fiat currencies.

We know that fiat currencies have inflation or deflation, and cryptocurrencies do not have such a problem, but, currency exchanges allow transaction between cryptocurrencies and fiat currencies. This means, that cryptocurrencies may break the supply-demand-balance in the fiat currency market. Moreover, the rate of trading pairs in each exchange makes difference to other exchanges, and such difference will generates arbitrage opportunities and triangular trading opportunities. This means, the merchant will buy Bitcoins on exchanges with lower rate and sell them on exchanges with higher rate, so they can earn money from the difference of rate, and these opportunities may consequently harass the balance in financial market and even exchange rate between two fiat currencies. Here arises a question whether cryptocurrencies replace fiat currencies or coexist with each other.

According to the present situation, the possibility that cryptocurrencies will replace fiat currencies will hardly appear. At first, cryptocurrencies have limitation, for instance, there are only 21 million Bitcoins. Secondly, cryptocurrencies are technology-based[15], they must be mined, so people have pretty big requirement on technology. Cryptocurrencies are complex[15], not only due to their algorithmics, but also because most people do not really understand how they work. Besides, cryptocurrencies are nowadays not pure currencies any more. Merchants treat them as well as financial assets. Merchants prefer to invest than to use. In the case of Bitcoin, there are totally about 13.5 million Bitcoins have already been

doi: 10.2313/NET-2015-03-1_13

mined, but there are now only about 200,000 Bitcoin in transaction each day in December (Figure 6), the transaction volume is very low. Merchants treat cryptocurrency as financial assets more than as currency, which makes cryptocurrency seem to be like gold or silver, so that cryptocurrency will slowly lose its function as currency.

At present, Cryptocurrencies provide a new possibility for people to do their business and a new lifestyle, an easy and efficient way, but it will not replace fiat currencies, at least in a short period will not.

Coexistence is no doubt the best consequence of the battle between these two completely different kinds of currencies, both of them can play to their strengths.

Cryptocurrency can optimize fiat currency[7], this means, cryptocurrency can fix the shortcoming of fiat currency. For example, with the help of cryptocurrency, people are able to transfer their money abroad, they do not need to worry about any extra condition of banks. In another word, it will be more efficient and easier if banks can provide services such as free abroad transaction.

Fiat currency can optimize cryptocurrency as well[7]. With the help of fiat currency, the acceptance of cryptocurrency will increase more quickly. The volatility of cryptocurrency will also be improved , so that people can concern more with payment for goods and services than speculative investment.

Nothing is impossible, according to the coexistence of cryptocurrencies and fiat currencies, we look forward to a new world in a hybrid economy, where customers and merchants alike must be able to get into and out of cryptocurrency quickly, easily and securely.

3.2.3 Cryptocurrency Security

As we learn about cryptocurrency, we will find that the security is one of the most significant features. As I mentioned before, firstly, cryptocurrencies such like Bitcoin verify transactions with the same state-of-the-art encryption that is used in military and government applications. Secondly, all transactions are anonymous, nobody can get merchant's real personal information or even steal their coins in the account, and that's why theoretically the merchants should not pay attention to their cryptocurrency account. But, cryptocurrencies is still not mature enough. It is not perfect, so some people still try to break the transaction system of cryptocurrencies exchange in order to steal coins from other people. It was revealed in February 2014 that more than $350 million worth of Bitcoins were stolen from Mt.Gox which led to the shutdown of this exchange! After this event, the price of Bitcoin felt from more than 800$ in January down to 577.52$ at the end of February. Because of this event, the merchants of cryptocurrencies reconsider about the security of cryptocurrencies and exchanges, and exactly since then people begun to treat Bitcoin as assets more than as currencies. The Mt.Gox incident led to loss of merchant's confidence in cryptocurrencies, and such serious consequences indicate that the Mt.Gox incident caused irreparable harm to the credibility of the entire Bitcoin system.

All of above, the security of cryptocurrency is not so stable as people thought before, it still needs to be improved. It seems that cryptocurrencies slowly lose their strengths compare to fiat currency and this is also one of the reasons why cryptocurrency cannot really replace fiat currency.

4. COMPETITION IN CURRENCY EXCHANGE

Bitcoin is the largest cryptocurrency in market capitalisation, volume, acceptance and notoriety, but it's not the most valuable coin. Litecoin takes the second place after Bitcoin[1]. But since the Mt.Gox incident and the shutdown of Silkroad, the prospect of Bitcoin does not seem to be very optimistic. Many alternatives to Bitcoin suddenly appeared in the meantime and several new exchanges entered the market at the same time as well, hence the competition not only between cryptocurrencies but also between currency exchanges became much more intense than ever before.

Until mid-2013, Mt.Gox was the dominant exchange and by November 2013, it was still an important player with 27% of the trade, but it is no longer a dominant exchange. That means some other exchanges sprung up, for example, BTCChina, with 35% of the trades, Bitstamp, with 24% of the trade, and BTC-e, with 14% of the trade[2]. Because cryptocurrencies can be exchanged into different fiat currencies on each different exchange, apart from competition between different cryptocurrencies, there is also competition between different fiat currencies, and sometimes it seems that merchants may have arbitrage opportunities due to such differences.

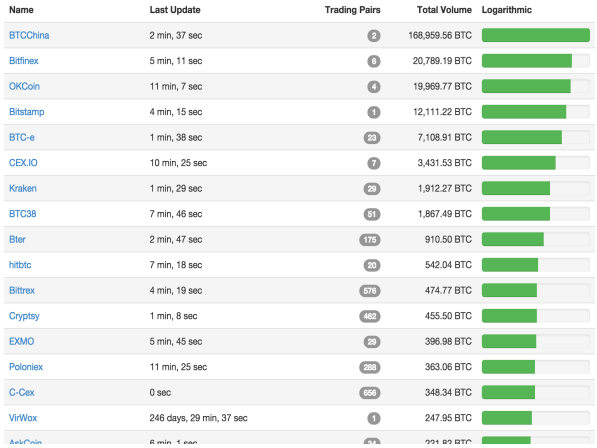
In this chapter, we are mainly concerned with the competition between different exchanges and the competition between cryptocurrency wallets and bank.

4.1 Competition among Different Exchanges

As I mentioned before, since Mt.Gox ceased operations, in large part due to a security breach — and a huge loss of Bitcoins, a large number of new exchanges appeared in the market and a large percentage of trade occurred at those new exchanges like BTCChina, OKCoin. In this chapter, we look at the competition among different exchanges after the Mt.Gox incident as well as factors which may influence such competition. Here, we only observe the data on 10.12.2014.

4.1.1 Competition in the Bitcoin Market

In the case of trades involving BTC, the following figure[9]



Name	Last Update	Trading Pairs	Total Volume	Logarithmic
BTCChina	2 min, 37 sec	2	168,959.56 BTC	
Bitfinex	5 min, 11 sec	6	20,789.19 BTC	
OKCoin	11 min, 7 sec	4	19,969.77 BTC	
Bitstamp	4 min, 15 sec	1	12,111.22 BTC	
BTC-e	1 min, 38 sec	23	7,108.91 BTC	
CEX.IO	10 min, 25 sec	7	3,431.53 BTC	
Kraken	1 min, 29 sec	29	1,912.27 BTC	
BTC38	7 min, 46 sec	51	1,867.49 BTC	
Bter	2 min, 47 sec	175	910.50 BTC	
hitbtc	7 min, 18 sec	20	542.04 BTC	
Bittrex	4 min, 19 sec	576	474.77 BTC	
Cryptsy	1 min, 8 sec	462	455.50 BTC	
EXMO	5 min, 45 sec	29	396.98 BTC	
Poloniex	11 min, 25 sec	286	363.06 BTC	
C-Cex	0 sec	656	348.34 BTC	
VirWox	246 days, 29 min, 37 sec	1	247.95 BTC	

Figure 8. Cryptocurrency Exchange / Markets List (Accessed on 10.12.2014) (Source: http://www.cryptocoincharts.info/markets/info).

doi: 10.2313/NET-2015-03-1_13

shows the current ranking in trades of Bitcoin. We can see that in 24 hours, BTCChina gains the most total volume with 168,959.56 BTC, Bitfinex takes the second place with 20,789.19 BTC, the third is OKCoin with 19,969.77 BTC, the fourth and the fifth are Bitstamp with 12,111.22 BTC and BTC-e with 7,108.91 BTC. These five exchanges hold the most current Bitcoin capitalisation in the market, the other exchanges are active in the market as well, but the volume traded is extremely small. That means, the competition is mainly occurred among these five exchanges. Because each of these exchanges has different trading pairs, it is difficult to compare these exchanges very exactly. In the following part of this subchapter we will consider more details of cryptocurrency exchanges within one trading pair.

In the case of trades involving BTC and the USD (Figure 3), by mid-December 2014 there are three major exchanges: Bitfinex, OKCoin, and Bitstamp. Bitfinex has about 32.52% of the volume and in value of $ 7,130,960 for the currency pair BTC/USD, while OKCoin has about 24.59% of the volume in value of $ 5,391,550 and Bitstamp has about 19.78% of the volume in value of $ 4,336,740 for this currency pair. These three current major exchanges occupy more than 70% of the total volume in trades of BTC and the USD, and their total trading volume reaches the worth of $ 16,859,250. Additionally, if we look at the ranking by mid-February 2014, BTC-e was the first of three major exchanges and Bitfinex took only the third place at that time, but now BTC-e takes only the fourth place on the ranking with 8.70% of the volume. Such a quick change indicates that the cryptocurrency market is always keeping changing, the strength and potential of each exchange should not be underestimated no matter how extremely small its volume was treaded before. The competition among three major exchanges is still very intense due to their tiny difference of volume.

Apart from the USD, the only global currency, Chinese yuan also takes an important role in the financial market nowadays. Especially after the Mt.Gox incident, a large percentage of trade occurred at two exchanges in China: BTCChina and OKCoin. BTCChina and OKCoin are two major cryptocurrency exchanges in trade of BTC/CNY, and there are still a little difference between BTCChina and OKCoin, BTCChina only allows the Chinese yuan as the only fiat currency, but OKCoin allows both Chinese yuan and US dollar as fiat currencies. In Chinese market, on December 2013 the People's Bank of China announced banning financial institutions from processing transactions in Bitcoin, the once-dominant BTCChina lost a lot of volume and the price of Bitcoin fell by more than half, in contrast, the initially small exchange OKCoin gained a lot of volume[10] because it allows the transaction with USD. Even though the BTCChina is still the largest Bitcoin exchange in China, and OKCoin is still the most powerful competitor. In the recent 24 hours (on 10.12.2014), BTCChina has already treaded 168,959.56 BTC but OKCoin has very little trading volume for BTC/CNY.

At last, the case of BTC/EUR should be considered about as well, although Euro does not play such an important role as US dollar and Chinese yuan plays, it still has a significant percentage of trade at most exchanges. Comparing to US dollar and Chinese yuan, Euro has a relative small percentage in the Bitcoin market, the largest exchange which allows Euro as fiat currency in trade is Kraken, and it has about 3.28 % of the volume and in worth of $ 600,329 for currency pair BTC/EUR. The second largest exchange in trade of BTC/EUR is itBit with only about 0.23% volume for this currency pair. The competition for trading pair BTC/EUR seems less intense than for other two major trading pairs.

4.1.2 Competition in the Litecoin Market

The data of Bitcoin market alone may be not convincing enough, so in this subchapter we will look at the data of Litecoin market as a represent market of altcoin market.

In the Litecoin market, there is much less trading volume than in the Bitcoin market. Among those exchanges (Figure 4), we observe that OKCoin has the largest trading volume with $ 2,047,790 for the trading pair LTC/USD and has about 70.83% of the total volume in the Litecoin market. The following exchange is BTC-e. which has $ 287,866 worth trading volume, and it has only about 9.96% of the total volume in the Litecoin market. Obviously, OKCoin has an absolute position in the LItecoin market. I think there will not be any other exchanges in the Litecoin market which could surpass OKCoin temporarily.

4.1.3 Influence Factors in the Competition among Exchanges

Through all the data above, we can find that OKCoin has a very strong competitiveness in both Bitcoin market and Litecoin market, and the major exchanges have usually more than 70% of the volume or even more in each cryptocurrency market. This means the difference of trading volume between major exchanges and other exchanges is quite large. Given these data and discoveries, the question arises why the difference is so large or what the reason is.

In my opinion, the first reason may be the motivation of arbitrage, the profitable opportunities from trading the same pair of currencies on two different exchanges, because the prices of the same cryptocurrency on different exchanges is different. There are tests for trading opportunities across exchanges[2], which examine potential trades involving USD/BTC and compare the exchange rate between the USD and BTC on BTC-e and Bitstamp, the former largest exchange trading BTC/USD. The result of these tests shows that on most days, Bitcoin was cheaper on BTC-e than on Bitstamp, and on half of the days, the difference in prices would yield more than 2% gain[2]. Apart from this result, the data of these tests also shows that for 5% of the day the prices at midnight on these two exchanges were different by more than 4%[2]. These tests examine trading opportunities for Litecoin as well. Finally, all the data suggest that gross trading opportunities were much greater across exchanges than within exchanges.

The second reason may due to the presence of "winner take all" effect. This means, the more volume traded within one exchange, the more safer and reliable this exchange will be, and the more merchants will exchange their currencies on this exchange.

Besides these two reasons which refer to the market itself, there are also some causes due to intervention of government. The reduction of Bitcoin trading volume on BTCChina on 5 December 2013 is an example of such a case. Although cryptocurrencies are decentralized currencies, the attitude of government could still have impact on their trading volume and

doi: 10.2313/NET-2015-03-1_13

even their prospective development. This kind of situation may be more obvious and easier in China, because there's just one government entity to deal with rather than the collection of state and federal agencies in the U.S.

Taken all, different price, different security and different reaction of government may be three main factors which have the most impact on currency exchanges and the competition among them.

4.1.4 Prospect of Cryptocurrency Exchanges

All in all, among these exchanges, OKCoin seems to be the most competitive cryptocurrency exchange not only in the US dollar market but also in the Chinese yuan market. It's high percentage of volume in both markets ensures its top position in cryptocurrency market. From my point of view, OKCoin will be a leader among cryptocurrency exchanges, at least in a short-term future.

After observing different performances of cryptocurrency exchanges, most cryptocurrency exchanges put much more focus on US dollar market than on any other currency markets. There is no doubt that US dollar is the most powerful fiat currency in cryptocurrency market, but focusing only on US dollar market may restrict the expand of cryptocurrencies and cause potential loss of users and opportunities, so I think some exchanges should allow more fiat currencies in transaction and this will be also helpful to expand acceptance and to gain more popularity.

Additionally, many exchanges have a limited number of cryptocurrencies which are allowed in trade. Maybe it benefits increase in volume of transaction of target cryptocurrencies or it may be easier to control all the transaction within this exchange, but more currencies bring more opportunities.

However, no matter how many currencies are allowed in transaction within each exchange, from the perspective of development, I think there are two main points which should be concerned with: security and popularity. Since Paypal and Microsoft build partnerships with Bitpay and Coinbase, cryptocurrencies have here a giant leap into the mainstream financial market and therefore have a better prospect than ever before.

With the support from enterprises and cryptocurrency merchants, with a better improvement in security, cryptocurrency exchange will find its new position in financial market.

4.2 Competition Between Digital Wallet and Bank

Once cryptocurrencies appeared in the market, some medias released articles which expressed the doubt about cryptocurrency exchanges or the worry about the future of bank. In this section, we investigate the competition between cryptocurrency wallets and banks. We will compare the advantages and disadvantages of both cryptocurrency wallets and banks, especially the risks in cryptocurrency wallets.

4.2.1 Digital Wallet or Bank?

What is digital wallet? There is on exact definition of digital wallet, but on the webpage www.blockchain.info/wallet there is a vivid subtitle of digital wallet: Be your own bank. As the name suggests, digital wallet is another kind of bank. Merchants and sellers can trade their currencies using this wallet. At this moment, there are two different kinds of wallet, the first one is a local wallet, in which people can store their coins locally on the computer or template, for example, Bitcoin Core. The second one is an online wallet where people can store their coins online, and such kind of wallet is more like online banking, for example, Green Address. The current biggest Bitcoin wallet is coinbase. Due to the large amount of cryptocurrencies, we only observe Bitcoin wallet in this section.

As an emerge product, Bitcoin wallet has several advantages which attract a lot of cryptocurrency merchants and sellers. At first, there is no additional transaction fee using Bitcoin wallet, this means using Bitcoin wallet can minimize the extra fee which is usually paid to bank. Secondly, transaction using Bitcoin wallet is very quick and easy, for example, people can send Bitcoins via email, SMS and even Facebook. Thirdly, it is safe to use wallet, because people can store their coins locally so that your wallet won't be stolen easily. Fourthly, using Bitcoin wallet enable transactions world wide, a borderless transaction. There is no reversible transaction using Bitcoin wallet.

As the transaction volume of Bitcoin is getting more and the use range of Bitcoin is getting wider, users can purchase most things such as books, foods, games, servers with cryptocurrencies. It seems that Bitcoin has a optimistic prospect, but it still has several flaws and some of them are even so fatal, that Bitcoin wallet does not have enough power to compete with banks.

The question, about which the users of Bitcoin wallet care at most is no doubt the security of Bitcoin wallet. Today, the security of Bitcoin wallet is still further less than banks. Banks have nowadays rather stable, safe and mature online banking system, but the online Bitcoin wallet is still in infancy. On 2 July 2013, an online Bitcoin wallet named Input.io was stolen by a hacker, and it lost about 4100 Bitcoin in worth of $ 1,4 millions. Analogous also happened several times on other online Bitcoin wallets. Recently, after a system update of the world's largest Bitcoin exchanges and Bitcoin wallet provider — Blockchain.info lost some of their client's wallets and information[12]. In one word, Bitcoin wallet users have to continue concerning with their security risks.

Of course, there are some others factors which limit the competitiveness of Bitcoin wallets as well, for example, the reaction of government, and new changes in the financial market. A few days ago, one of the most popular ways for users to buy and sell bitcoins — localbitcoin.com has announced on Twitter that it will be no longer available in Germany due to regulatory reason[13].

Besides, the acceptance of Bitcoin has also impact on Bitcoin wallet, in the initial period, the acceptance of Bitcoin was very low, hence the Bitcoin wallet had hardly power, but now this situation is turning better, Bitcoin gains more popularity today and even Paypal, one of the biggest currency transaction system accepts Bitcoin. Since this September, Paypal took its first venture into the world of all-digital money, this means merchants can now start accepting Bitcoin as payment[11], and I think, such a behaviour of Paypal may

accelerate the acceptance of Bitcoin, especially on eBay, that could be also helpful for Bitcoin wallet.

In comparison with Bitcoin wallet, bank may be threatened by Bitcoin wallet but it still has absolute advantage against Bitcoin wallet, like its absolute high acceptance, its security, and people can get interests when they put their money in bank. Merchants can do much more business through banks than using Bitcoin wallet.

Unfortunately, since a serious financial crises, the trust between people and banks becomes weaker, some people do not trust central bank any more, in another word, the crisis of confidence is the most urgent issue which banks are facing ~~now~~. Such a crisis can help cryptocurrencies and digital wallet to get more benefits and to gain more popularity.

Generally, the advantage of banks is much greater than the disadvantage of banks, people are used to own a bank account or even more, and it is too hard to change the way people used to behave.

All of above, we observe that the competition between Bitcoin wallet and bank is still not so obvious or intense currently, the number of users is growing but still much lower than the number of bank account holders, and the development of digital wallet is surely being influenced by the development of cryptocurrencies. At the same time, banks have also urgent issues which need to be solved. The world is changing, thus the trend of Bitcoin wallet should not be underestimated by banks.

5. CONCLUSION

After a five-year-development, cryptocurrency has experienced ups and downs, it may have a large long-term effect on both currency and payment systems or bank, but these cryptocurrencies are currently still in their infancy, there are still a lot improvements need to be done, and there are still lots of difficulties which have to be overcame.

There are two important periods within the development of Bitcoin: May-September 2013 and October 2013-February 2014. In the first period, Bitcoin's price was relatively stable, while in the second period it was very volatile.

The anonymity and P2P transaction brings benefits as well as troubles, although cryptocurrency is decentralized, the regulatory of government influenced its development and the competition against fiat currency. In the view of the existing condition, Bitcoin maintains the dominant position in the competition against other cryptocurrencies, but if it compares to fiat currency, Bitcoin has has only a few power. Cryptocurrency can not replace fiat currency right now, and it does not need to replace currency too. When money becomes data, the better solution for both cryptocurrency and fiat currency should be hybrid economy, that both kinds of currencies can play to their strengths and optimize each other.

In the competition among cryptocurrency exchanges, OKCoin has a top position not only in the Bitcoin market but also in the Litecoin market. China is now one of the biggest market for cryptocurrency, the transaction of BTCChina is the best prove. Although US dollar has a dominant position in trade with cryptocurrency, the potential of other currencies like Chinese yuan and euro should not be underestimated. In addition, through the comparison of digital wallet and bank, we observe that the digital wallet is still not very competitive against bank or other payment systems, it has potential to improve and to make progresses.

Clearly, cryptocurrency and its exchange have optimistic prospect yet. As more and more payment systems such as Paypal support the payment with cryptocurrencies, and if cryptocurrency can make itself more user-friendly, it will have a better development, and as the adoption of cryptocurrency increases, its volatility may have significant improvement. In the long-term, it is not sure whether the advantages of cryptocurrency will be sufficient to gain more volume of the financial market and have a good cooperation with fiat currency and payment systems, but we look forward that someday it is able to change and improve our lifestyle as a model for other technology innovations.

6. REFERENCES

[1] Greydon, Carter. September 16, 2014. What is Cryptocurrency. DOI= https://www.cryptocoinsnews.com/cryptocurrency/

[2] Gandal, N., Halaburda, H. September 6, 2014. Competition in the Cryptocurrency Market. DOI= http://weis2014.econinfosec.org/papers.

[3] DOI=http://www.cryptocoincharts.info

[4] 451 Research's 2014 US Consumer Survey, June.

[5] DOI= http://en.wikipedia.org/wiki/Cryptocurrency

[6] Grennberg, Andy. April 20, 2011. Crypto Currency. DOI= http://www.forbes.com/forbes/2011/0509/technology-psilocybin-bitcoins-gavin-andresen-crypto-currency.html.

[7] Neville, Sean. March 9, 2014. Forget Bitcoin vs Fiat, Welcome the Hybrid Economy. DOI= http://www.coindesk.com/forget-bitcoin-vs-fiat-welcome-hybrid-economy/.

[8] DOI= http://www.cryptocoincharts.info/markets/info.

[9] DOI= http://coinmarketcap.com/currencies/volume/24-hour/.

[10] Philips, Matthew. March 20, 2014. Bitcoin Isn't Banned in China—and It's Quickly Gaining Ground. Program. DOI= http://www.businessweek.com/articles/2014-03-20/btc-chinas-bobby-lee-Bitcoin-isnt-really-banned-in-china-and-its-quickly-gaining-ground.

[11] Pagliery, Jose. September 26, 2014. Paypal now lets shops accept Bitcoin. DOI= http://money.cnn.com/2014/09/26/technology/paypal-bitcoin/.

[12] December 15, 2014. Blockchain.info loses Bitcoins stored in "a few hundred addresses" during software update. DOI= http://bitcoinexaminer.org/blockchain-loses-bitcoins-few-hundred-addresses-software-update/.

[13] Carlouro, Eric. September 8, 2014. LocalBitcoin.com: Servercies Will No Longer Be Available in Germany. DOI= http://newsbtc.com/2014/12/08/localbitcoins-com-services-will-no-longer-available-germany/.

[14] DOI= https://bitcoin.org/en/faq#what-are-the-advantages-of-bitcoin.

[15] Delono, John. November 9, 2013. Bitcoin Simply Canoont Replace Fiat. DOI= http://letstalkbitcoin.com/bitcoin-simply-cannot-replace-fiat/.

ISBN 978-3-937201-47-4
DOI 10.2313/NET-2015-03-1

ISSN 1868-2634 (print)
ISSN 1868-2642 (electronic)