University of Zurich[UZH]

Burkhard Stiller, Christos Tsiaras, Andri Lareida, Lisa Kristiana, Thomas Bocek, Patrick Poullie, and Corinna Schmitt (Eds.)

Internet Economics X

TECHNICAL REPORT – No. IFI-2016.01

January 2016

University of Zurich
Department of Informatics (IFI)
Binzmühlestrasse 14, CH-8050 Zürich, Switzerland

ifi

B. Stiller et al. (Eds.): Internet Economics X
Technical Report No. IFI-2016.01, January 2016
Communication Systems Group (CSG)
Department of Informatics (IFI)
University of Zurich
Binzmühlestrasse 14, CH-8050 Zürich, Switzerland
URL: http://www.csg.uzh.ch/

Introduction

The Department of Informatics (IFI) of the University of Zürich, Switzerland works on research and teaching in the area of communication systems. One of the driving topics in applying communications technology is addressing investigations of their use and application under economic constraints and technical optimization measures. Therefore, during the autumn term HS 2015 a new instance of the Internet Economics seminar has been prepared and students as well as supervisors worked on this topic.
Even today, Internet Economics are run rarely as a teaching unit. This observation seems to be a little in contrast to the fact that research on Internet Economics has been established as an important area in the center of technology and economics on networked environments. After some careful investigations it can be found that during the last ten years, the underlying communication technology applied for the Internet and the way electronic business transactions are performed on top of the network have changed. Although, a variety of support functionality has been developed for the Internet case, the core functionality of delivering data, bits, and bytes remained unchanged. Nevertheless, changes and updates occur with respect to the use, the application area, and the technology itself. Therefore, another review of a selected number of topics has been undertaken.

Content

This new edition of the seminar entitled "Internet Economics X" discusses a number of selected topics in the area of Internet Economics. the first talk "Security Policy Reconfiguration Solutions in Wireless Sensor Networks" discusses security-related aspects of Wireless Sensor Networks with a focus on the request for policy reconfiguration. Talk two "Private Clouds: Technology and Deployment" discusses the various aspects that need to be taken into consideration when creating private clouds raising from financial aspects to security issues. Talk three "VANET Services on the Road: A Contrary Perspective of VANET Services" introduces a variety of aspects that need to be addressed when Vehicular Ad Hoc Network (VANET) services should be developed and established to fulfill the user's request to be online all the time even on the road. Talk four on "Socioeconomic Impact of Missing Parking Availability Information" presents a solution addressing parking problem I'm crowded and for the driver unknown areas. Nowadays watermarking is nearly everywhere embedded that finds now the way to sensor data sets, which is characterized in detail in talk five on "Watermarking in Sensor Data Sets". Talk six on "Applying Social Network Analysis (SNA) to P2P File" presents technological and economic aspects of applying SNA to P2P files. Talk seven "How to Get Users to Use CoinBlesk" discusses the new paying option with digital currency. Finally, the talk "Certificates" provides a detailed characterization of certificate generation and usage, especially of X.509 certificates.

Seminar Operation

Based on well-developed experiences of former seminars, held in different academic environments, all interested students worked on an initially offered set of papers and book chapters. Those relate to the topic titles as presented in the Table of Content below. They prepared a written essay as a clearly focused presentation, an evaluation, and a summary of those topics. Each of these essays is included in this technical report as a separate section and allows for an overview on important areas of concern, sometimes business models in operation, and problems encountered.

In addition, every group of students prepared a slide presentation of approximately 45 minutes to present his findings and summaries to the audience of students attending the seminar and other interested students, research assistants, and professors. Following a general question and answer phase, a student-lead discussion debated open issues and critical statements with the audience.

Local IFI support for preparing talks, reports, and their preparation by students had been granted by Lisa Kristiana, Patrick Poullie, Andri Lareida, Thomas Bocek, Christos Tsiaras, Corinna Schmitt, and Burkhard Stiller. In particular, many thanks are addressed to Corinna Schmitt for her strong commitment on getting this technical report ready and quickly published. A larger number of pre-presentation discussions have provided valuable insights in the emerging and moving field of Internet Economics, both for all students and supervisors. Many thanks to all people contributing to the success of this event, which has happened in a lively group of highly motivated and technically qualified students and people.

Zürich, January 2016

Contents

1 Security Policy Reconfiguration Solutions in Wireless Sensor Networks **7**
Anastasia Ruvimova, Raphael Ochsenbein, Sanjiv Jha

2 Private Cloud: Technology and Deployment **31**
Geetha Parangi, Linda Diep, Elfat Esati

3 VANET Services on the Road: A Contrary Perspective of VANET Services **53**
Arlind Bejta, Jens Hinkelmann, Sascha Gabler

4 Socioeconomic Impact of Missing Parking Availability Information **73**
Adriano Meyer Broyn, Stefan Bublitz, Sacha Uhlmann

5 Watermarking in Sensor Data Sets **97**
Melina Mast, Romana Pernischova, Manuel Schlegel

6 Applying Social Network Analysis (SNA) to P2P File Sharing **123**
Andreas Schaufelbühl, Robin Stohler, Benjamin Bürgisser

7 How to Get Users to Use CoinBlesk **149**
Cyrus Einsele, Mervin Cheok, Patrick Widmer

8 Certificates **169**
Noah Zani, Tim Strasser, Andrés Baumeler

Chapter 1

Security Policy Reconfiguration Solutions in Wireless Sensor Networks

Anastasia Ruvimova, Raphael Ochsenbein, Sanjiv Jha

With the advancement of wireless technology, Wireless Sensor Networks (WSNs) are gaining popularity in a variety of areas, from healthcare to wildfire surveillance. Unfortunately, these devices tend to have very limited storage, energy, and processing capabilities, and are therefore not able to support complex, high-overhead security mechanisms. In addition, these networks are often situated in remote environments and are therefore especially vulnerable to malicious attacks. Each node in the WSNs are given some security mechanisms to work in different enviroment, now the question is "How do we choose a security mechanism with low enough cost which will protect the network from as many threats as possible?". In this paper, we analyze three dynamic reconfiguration solutions for WSNs which will allow the nodes to optimize their security mechanisms by adapting to their current situation and needs.

Contents

1.1 Introduction **9**
1.2 Security Considerations in WSNs **11**
1.2.1 Security Vulnerabilities in WSNs 11
1.2.2 Defense Strategies for WSNs 14
1.3 Security Policy Reconfiguration Solutions in WSNs **18**
1.3.1 Effective Key Management in Dynamic Sensor Networks 18
1.3.2 Trust-based Enforcement of Security Policies 19
1.3.3 Dynamic Reconfiguration - Inter-trust & Famiware 24
1.3.4 Evaluation of Inter-trust and FamiWare 27
1.4 Summary and Conclusion **28**

1.1 Introduction

A Wireless Sensor Network (WSN) is a collection of distributed autonomous sensors to monitor physical or environmental conditions, such as temperature, sound, pressure, gases, and to cooperatively pass their data through a wireless network to a decision maker's location. For example, the Intelligent Transport System (ITS) is one of the most relevant application today, where the security issues in WSNs play a very important role. These systems have the potential to improve mobility, safety and security, while ensuring energy efficiency and reducing the environmental impact of transportation systems.

These networks consist of a host of sensors and many different types of sensors are being used for different scenarios. These include military, traffic control, healthcare, environmental monitoring, agriculture, and more. The implementation of WSNs has increased in number, as have the security attacks. To make the network secure and reliable, different security policies need to be implemented. To make network more robust and achieve required security, the network is composed of many different types of sensors. As the heterogeneity of sensors increases, the security requirement increases at the node level. The different types of nodes are constrained by the node architecture, including the operating system (OS), the memory, and the processing power. Thus, generic security policies are not an ideal solution, as almost no two WSNs have exactly the same purpose and requirements.

To reuse the example of ITS again: they are composed of different kinds of devices installed at several locations, such as sensors inside the vehicles and sensors on the traffic signals. For instance, the vehicles can be equipped with several sensors with different sensor units, such as acoustic, accelerometers, temperature, the vehicle's on-board unit (OBU) computers, and global positioning system (GPS) devices. These sensors are used to ensure safe driving and responsibility: for instance, to measure the distance from the obstacles nearby. The vehicles also use information received from surrounding vehicles, such as their speed or distance, and are able to communicate with each other in order to cooperate for safety and other reasons. Furthermore, the traffic signals send information to the vehicles informing them of speed limits and general traffic regulations. ITS have a wide variety of interesting applications, such as road safety applications.

Using a large number of sensors increases the chances of communication disturbance and, in some cases, this information is highly confidential (e.g., research work satellite communication). The security needs of all of these applications are quite significant.

In case of ITS, to know whether a node in the network is being uncooperative or malicious is really crucial for preventing intentionally wrong instructions being sent to a vehicle. Furthermore, these systems are very dynamic, and their contextual environment changes very frequently, so the security policies need to adapt to the changing environment. The ITS infrastructure is composed of extremely heterogeneous devices where both the monitoring and the deployment and management of the security policies must be performed in different ways.

In ITS, the WSN is used to indicate that driving at the recommended speed with a traffic congestion in the vicinity, could cause an accident (shown in Figure 1.1). Therefore, the messages sent between the nodes and the base station must be properly encrypted, and all of the nodes must be trustworthy. A protection of the whole system is required. However, very strict security policies have a penalty in efficiency, especially in energy and performance - two typical requirements of these applications. Therefore, the security policy to be applied in each situation must be negotiated to make a balance between the necessary security and the efficiency. Additionally, as these situations can change quickly, these policies have to be renegotiated constantly. For instance, in case a malicious software is suspected in a given WSN (detected by intruder detection algorithms), the security needs to be increased, and the suspect nodes must be isolated. To the contrary, on roads

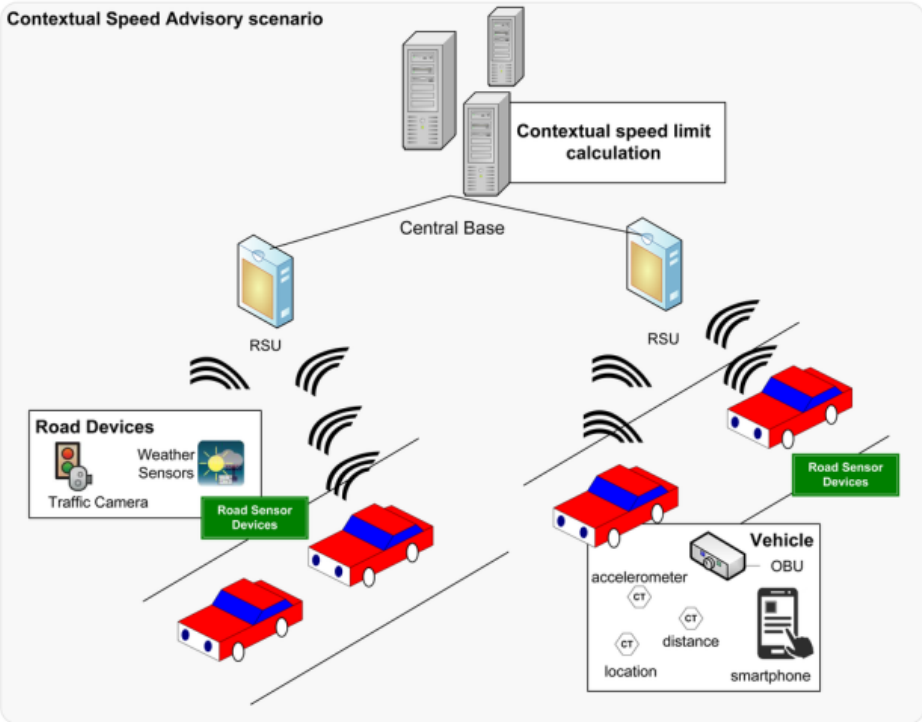


Figure 1.1: A context speed advisory (CSA) scenario [10].

where the main interest is to preserve the lifetime of the system and where there is no risk of malicious attacks, the security can be a little bit weaker. Similarly, in situations where communication performance is required (e.g., high traffic density) the security could also be relaxed. On the other hand, in exceptional circumstances, the interception of data sent by certain vehicles (e.g., a presidential motorcade) has to be totally avoided. In this case, the security policy applied to these specific nodes must be modified.

There is some important effort being made to develop specific algorithm for WSNs in the fields of trust, security, and privacy, such as new algorithms for authentication, positioning, network intrusion detection, efficient key management for dynamic Wireless sensor networks. As a network has many different sensors with different requirements, if the same algorithm has been implemented in all the nodes and the working environment has changed, the present algorithms for some specific nodes needs improvement - i.e., new algorithms are to be implemented within the nodes to work in the changed environment, which was not mainly done by the previous algorithm. So, the nodes have to be adaptive in order to change accordingly. The main issues captured in the new advanced algorithm are low-level programming abstractions, limited capabilities, and different routing protocols. The sensor nodes need to be adaptive to environmental change with the lowest possible impact for users and without stopping the devices.

The rest of the paper is organized as follows. Section 1.2 discusses different security considerations in WSNs with the security vulnerabilities and defense strategies for WSNs. Known threats and vulnerabilities of WSNs are introduced, and the theoretical solutions to them presented. Section 1.3 presents three practical solutions for security re-configuration in WSNs. Here, we discuss three approaches for dynamic reconfiguration of security policies in WSNs. In section 1.4 we conclude our paper.

1.2 Security Considerations in WSNs

In this section we analyze known Denial of Service (DoS) threats and attacks on privacy and authentication on the four primary levels: the physical layer, link layer, network layer, and transport layer.

1.2.1 Security Vulnerabilities in WSNs

For many of the before mentioned WSN applications, security is of very high priority. WSNs are often deployed in remote locations and left unattended (for example, sensors which watch for fire activity in forests) and are, therefore, especially vulnerable to attacks. Unfortunately, the resource constraints of these sensors do not allow for traditional high-overhead security mechanisms. We will propose several solutions after analyzing the threats we need to protect against.
These attacks may be categorized into attacks on security and authentication and DoS attacks. Attacks on security and authentication include packet modification, eavesdropping, and spoofing of packets. As described by Sen, DoS attacks *"may permit real-world damage to the health and safety of people. The DoS attack usually refers to an adversary's attempt to disrupt, subvert, or destroy a network. However, a DoS attack can be any event that diminishes or eliminates a network's capacity to perform its expected functions"* [4].

1.2.1.1 DoS Attacks

DoS attacks in WSNs are sometimes very costly and have received much attention from researches. These attacks exist on four levels: the physical layer, the link layer, the network layer, and the transport layer. We will analyze the threats for each in the following.
The physical layer is responsible for *"frequency selection, carrier frequency generation, signal detection, modulation, and data encryption"* [4]. These attacks are especially threatening because WSNs may be deployed in hostile environments where an attacker has physical access. The two types of attacks on the physical layer are jamming and tampering. The first attack, jamming is defined as an attack which interferes with the node's radio frequencies used for communication. Even intermittent jamming may prove destructive as messages in network communication are often time-sensitive. A jamming source may disrupt the entire network.
Because sensor networks are often deployed in remote outdoor environments and left unattended, nodes are also highly vulnerable to tampering. Sen warns that *"physical attacks may cause irreversible damage to the nodes. The adversary can extract cryptographic keys from the captured node, tamper with its circuitry, modify the program codes or even replace it with a malicious sensor. It has been shown that sensor nodes such as MICA2 motes can be compromised in less than one minute time"* [4].
The next level up, the link layer, is responsible for multiplexing of data-streams, data frame detection, medium access control, and error control. This layer is vulnerable to attacks which include: resource exhaustion, purposefully created collisions, and unfairness in allocation [4]. A collision is defined as an attack where two nodes attempt to transmit on the same frequency simultaneously, causing the packets to be discarded and require retransmission. Sen elaborates on these attacks, noting that: *"an adversary may strategically cause collisions in specific packets such as ACK control messages. A possible result of such collisions is the costly exponential back-off. The adversary may simply violate the communication protocol and continuously transmit messages in an attempt to generate collisions"* [4].
Strategically-placed collisions may also result in resource exhaustion, for example: when a naive link layer node continuously attempts to retransmit the corrupted packets. The

third attack, unfairness of allocation is a weak form of DoS attack caused by intermittently employing collision and resource exhaustion: *"the adversary causes degradation of real-time applications running on other nodes by intermittently disrupting their frame transmissions"* [4].
Attacks on the network layer of WSNs include spoofed routing information, selective packet forwarding, sinkhole, Sybil, wormhole, hello flood, acknowledgment spoofing, and more. A brief description of these attacks follows.
The first, spoofed routing information directly targets the network's routing information by altering, spoofing, or replaying routing information to disrupt the network's traffic [4]. As Sen describes, these disruptions include *"creation of routing loops, attracting or repelling network traffic from selected nodes, extending or shortening source routes, generating fake error messages, causing network partitioning, and increasing end-to-end latency"* [4]. In a multi-hop network like a Wireless Sensor Network, all the nodes need to forward messages accurately for message communication.
An attacker may compromise this communication through selective forwarding, where some messages are forwarded and others are dropped. In a sinkhole attack, a node's routing information is altered to make it look more attractive than others. As a result, its neighbor nodes choose this compromised node for data routing.
In a sybil attack, a node is made to present multiple identities in the network, which disrupts the data distribution in peer-to-peer networks. Sen notes that *"in addition to defeating distributed data storage systems, the Sybil attack is also effective against routing algorithms, data aggregation, voting, fair resource allocation, and foiling misbehavior detection. [...] For example, may take on more than one identity in order to generate additional 'votes'. Similarly, to attack the routing protocol, [...] a malicious node taking on the identity of multiple nodes, and thus routing multiple paths through a single malicious node"* [4].
Our next attack is the wormhole - is a low latency link through which messages can be replayed between two parts of the network. This is similar to a sinkhole attack in that it creates a one-hop link to the base station through the attacking node in a far part of the network.
The hello flood attack abuses the common assumptions among Hello packet protocols that receiving a packet implies that the sender is within radio range of the receiver. The attacker can use a high-powered transmitter to fool many nodes into thinking they're within the neighborhood. As a result, the attacker broadcasts a shorter route to the base station and all the nodes which receive the Hello packet attempt to transmit to the attacker node, but the nodes are out of range.
Some algorithms for routing WSN's require acknowledgement packet transmission. We define acknowledgment spoofing as an attack where the malicious node overhears packet transmissions from its neighbor nodes and spoofs the acknowledgements, providing false data to the nodes. This results in false information being disseminated to the nodes.

1.2.1.2 Transport Layer

The Transport Layer is vulnerable to flooding attacks and desynchronization attacks. The flooding attack exhausts the node's memory by repeatedly making a new connection request until resources required by each connection have reached a maximum limit. Thus, further legitimate requests are ignored.
An attacker may use desynchronization by disrupting existing connections. This can be done by repeatedly spoofing messages to an end host, resulting in the request for retransmission of missed framed. The attacker may time this to hinder the end host's ability to exchange data, causing them to waste energy attempting to recover from errors which do not exist.

To summarize, we list the layers and their respective attack vulnerabilities in Table 1.2.1.2.

Table 1.1: Attack Vectors in the Network Layer

Layer	Attack
Physical Layer	Jamming, tampering
Link Layer	Collision, resource exhaustion, unfairness in allocation
Network Layer	Spoofed routing information, selective packet forwarding, sinkhole, Sybil, wormhole, hello flood, acknowledgement spoofing
Transport Layer	Flooding, de-synchronization

1.2.1.3 Attacks on Privacy

Let us now analyze attacks on secrecy and authentication. We begin with the node replication attack in which an attacker copies a node into the network by replicating an existing node's identifier. This node copy may cause severe damage to the communication by corrupting and forwarding the packets in wrong routes. Sen adds that *"this may also lead to network partitioning, communication of false sensor readings. In addition, if the attacker gains physical access to the entire network, it is possible for him to copy the cryptographic keys and use these keys for message communication from the replicated node. The attacker can also place the replicated node in strategic locations in the network so that he could easily manipulate a specific segment of the network, possibly causing a network partitioning"* [4].

In addition to DoS attacks, WSNs are vulnerable to attacks on privacy and authentication. This is partially a side effect of the network's ability to automate data collection through efficient sensor deployment. Unfortunately, protecting sensitive data in wireless networks is no easy task. This is especially so because seemingly innocuous data might give the attacker enough material to derive sensitive data. Moreover, a large amount of data is easily accessible through remote access mechanisms, so the attacker does not need to be present to get his or her hands on the data and may remain anonymous. The adversary may even monitor several sites at once. We will take a look at a few common attacks on sensor data privacy.

The simplest and most common attack on information privacy is eavesdropping and passive monitoring. Without cryptographic protection, the attacker can easily understand the contents. Sen warns, *"packets containing control information in a WSN convey more information than accessible through the location server. Eavesdropping on these messages prove more effective for an adversary"* [4]. For even more effectiveness, eavesdropping can be combined with traffic analysis. If done right, the adversary can pick out nodes which have special roles and activities in the network. For example, *"a sudden increase in message communication between certain nodes signifies that those nodes have some specific activities and events to monitor"* [4].

The attacker may use camouflage by compromising a senor node in the network and later using it to masquerade as a normal node. This false node may send false routing information and attract packets from other nodes to forward further. Having received the packets, the node forwards them to strategic nodes where privacy analysis is carried out systematically on the packets. As seen above, wireless sensor networks are subject to threats on many levels and by numerous approaches, making it even more difficult to find a universal security policy.

1.2.2 Defense Strategies for WSNs

In the previous section, a number of attack vectors on WSNs have been described. For most of these vectors, there exist various defense or at least mitigation mechanisms, which can in combination with each other be used to protect WSNs. However, due to the limited resources of the sensor nodes, the defense mechanisms have to be employed strategically so as to not overuse the computing power and run into run-time performance problems and to prevent overly consuming the battery power of the device.
The next section will focus on describing some select defense strategies for WSNs, namely those which are being used by the practical solutions for security policy reconfigurations introduced in section 1.3. Those attacks that have been detailed earlier, but are also not necessary in order to understand our proposed policy reconfiguration solutions are not described here.

1.2.2.1 Cryptography and Key Management

This performance dilemma may be most striking, when evaluating the cryptographic algorithms to be deployed in the network: Many researchers are hesitant in recommending public key cryptography for sensor networks, as the code and data size as well as the computational and energetic load are too high for sensor nodes [4]. One experiment shows that encrypting 1kB consumes 42mJ when using an RSA public key algorithm, whereas by using an advanced encryption standard (AES) encryption, slightly less than 1mJ is consumed [4]. Only recently researchers have published proof that some public key cryptographic algorithms are suitable for deployment in WSNs. While the optimizations that can be used to deploy public key cryptography in WSNs will be demonstrated in the practical section of this paper, one main limitation should be noted here: the focus has mainly been on optimizing the public key operation time, and relies on nodes with more computational power for the private key operations, which still remain expensive [4]. Therefore, these approaches might be less suitable for homogeneous sensor networks. Notably, the following algorithms have been shown to work in WSNs: Rabin's Scheme [13], Ntru-Encrypt [14], RSA [15], and Elliptic Curve Cryptography (ECC)[16] [17].
An alternative to public key cryptography is found in symmetric key cryptography, and, consequently, a lot of research focuses on these algorithms for the encryption of sensitive data in sensor networks. With these methods, a single shared key is used between the nodes talking to each other and one of the key challenges is the distribution of said key [4]. For this, it might be interesting to consider a hybrid approach, where a public key of the network key distribution center (KDC) is distributed to the sensor nodes, and the shared key for symmetric cryptography is distributed via public key cryptography.
A common approach in WSNs is to have the usually more powerful base station become the KDC for the network, as that usually requires keeping track of nodes and their keys, blacklisting keys in case a node corruption is detected [4]. However, the drawback of this approach is that the base station becomes a single point of failure for the network.
There are several approaches, with which a KDC can manage the network keys and encryption based on heuristics. As described earlier, one might distinguish between more sensitive traffic, such as symmetric key exchanges, which requires stronger encryption. In a deterministic approach, there are predefined encryption mechanisms for different packets. On the other hand, a probabilistic and distributed schemes might be chosen [4]. For instance, a random subset of a key ring can be shared with a number of nodes, which just have to find one common key to encrypt their communications. Furthermore, the subset of keys available to a node can be used as a fingerprint to identify a node [4].

1.2.2.2 Protecting Against Routing Protocol Attacks

Securing the routing protocol has the main goal of ensuring the integrity, authenticity, and correct transmission of messages in the network [4]. Consequently, this means securing all traffic in the network with one of the previously described encryption algorithms. There exist a number of secured routing protocols which can be used to secure the sensor network.
The first routing protocol is μTESLA (the micro version of the timed efficient streaming loss-tolerant authentication protocol), another protocol is INENS (intrusion tolerant routing protocol in WSNs), yet another protocol is SNEP (secure network encryption protocol), and there is also TRANS (trust routing for location aware sensor networks); additionally there are protocols built upon the other protocols, such as SPINS which incorporates μTESLA and SNEP [4]

1.2.2.3 Defense Against the Sybil Attack

In order to protect against the sybil attack, it is necessary that the nodes of the network can be correctly identified, and that every node exists only once. In other words, the network needs to be capable of identifying fake / multiplied nodes, for instance by keeping track of all the network nodes and their respective keys and identities. This can be achieved either by having specific keys for each node, or by assigning different radio channels to nodes in the neighborhood and listening to those channels [4]. In networks with random key predistribution, it is possible to use the keyring of a node to identify the node.

1.2.2.4 Managing the Node Replication Attack

Because centralized node identity management through the KDC or another central node has a single point of failure, and because distributed identity management protocols, such as neighborhood voting, would fail to detect distributed replications, collective identity verification algorithms have been brought forward. Both algorithms work by making a multicast to the network, which is then verified.
The first approach uses a randomized multicast of the node location to a randomly selected number of witnesses, whereby the number of witnesses has to be chosen to be large enough to guarantee that in each neighborhood there are at least two witnesses [4]. The number of witnesses can be calculated with the mathematics for the birthday paradox, which was originally used to calculate the probability of two people in a random group of people having the same birthday. Each node broadcasts its location to the network, and each node stores the location information of its neighbors. If a conflicting claim is found, the duplicate node is revoked [4]. This method can find all duplicates in the network, if all the broadcasts reach all the nodes in the network, however this broadcast then introduces a high total communication cost for the network.
The second method uses the network topology and a line-selected multicast to detect replication. In order to reduce the communication overhead of the first multicast approach, every node chooses one random node as a witness which will keep track of the node's location [4]. When the location is communicated to the witness, every node forwarding the information also keeps track of the location, and checks if there is any conflict. Therefore, if any node claims to be in an incorrect location, as soon as the two paths for the location witness intersect, it is possible to detect the fraudulent claim.

1.2.2.5 Preventing Traffic Analysis

If the sensor network has a single base station or key distribution center, then the network has a single point of failure. An attacker might want to identify as a crucial device in the network, and to do so, traffic analysis would be one option. To defend against such an attack, a multiple parent routing scheme should first be introduced, where any node can forward packets to one of many parents [4]. This creates a less deterministic network routing. Furthermore, the network hops towards the base station can be randomized through a controlled random walk. Additionally, random fake paths can be introduced, and fake communication hotspots or base stations can be added to the network. In combination, these methods together make it difficult to conduct traffic analysis for finding the base station.

1.2.2.6 Protecting Sensor Privacy

While encryption is one way to protect sensitive data, there is still a risk of the encrpytion being broken and the data being exposed. If the goal is to prevent the disclosure of sensitive data, even when encryption is broken, then it becomes necessary to strip identifying information from the data and make it anonymous.
The anonymization method that can be used to reduce the impact of a broken encryption key or a compromised node, is to store the sensitive data in a decentralized way in a tree structure over a range of nodes [4]. In order to do calculations on the data, a number of nodes has to be involved collaboratively. In addition to that, the methods for preventing traffic analysis and routing protocol attacks should also be employed.
Another method to protect data is to manage access control by specifying several policies that restrict data access, with parameters such as time of request, location, speed, and identity [4]. The drawback of using policies, however, is that they involve a server that enforces these policies, and therefore controls data access. Since it is impractical to enforce such policies on a node-level, again a single point for failure is introduced.
The next method for protecting sensitive data is information flooding. According to Sen, several flooding algorithms of varying sophistication can be used [4] The simplest approach is baseline flooding, where every node forwards the messages it receives only once, and never retransmits a message. To reduce the traffic, probabilistic flooding can be employed, where only a subset of nodes forward the data they receive. However, if an attacker intends to find the location of the source node, with the previous flooding algorithms, an attacker might be able to conduct traffic analysis in order to find the source node. Therefore, similar techniques as those preventing traffic analysis have to be employed for flooding as well. That means, fake messages have to be created, and the fake messages have to be created. However, here one does not only want to prevent finding the target of the traffic, but also the originator, and therefore phantom flooding should be used, which directs an attacker towards a fake source. To do so, first a random or directed walk to a phantom source is initiated, and the phantom source initiates a flooding request in order to deliver the message. Furthermore, the random walk has to be initiated from both source to sink, and from sink to source in order to prevent the collection of location information. To that end, the greedy random walk (GROW) protocol can be used.

1.2.2.7 Intrusion Detection

In a widely deployed sensor network, it must be assumed that some nodes can become compromised, especially so for low-power sensor nodes where less strict security mechanisms might be deployed due to computing and memory limitations of the node. For that reason, it becomes very important that the network has an intrusion detection system

(IDS) in place, which can monitor the network for suspicious activity patterns and find corrupted or compromised nodes in order to revoke their network access [4].
One approach is to have an IDS implemented at node level, where the nodes do not exchange intrusion data, whereas in a second approach, a cooperative IDS is implemented where each node has an intrusion detection agent to detect local attacks, but the intrusion data is exchanged between nodes in order to detect global attacks. A third approach is to use a hierarchical architecture, where nodes are clustered and cluster heads manage the local traffic and intrusion detection for the cluster [4]. One method to detect intrusion on a global is implemented with the interleaved hop-by-hop authentication (IHOP) scheme. In order to detect local intrusion, a local intrusion detection system such as local intrusion detection system (LIDS) can be deployed on each node [5]. LIDS can also work on a global level, if the nodes collaborate to exchange security data and intrusion alerts amongst each other [4].

1.2.2.8 Secure Data Aggregation

With the goal of reducing unnecessary data transmission in order to save energy, many sensor networks make use of data aggregation at a level of node clusters. However, if such an cluster head is compromised, it becomes easy to either collect sensitive information and to inject false data in the network [4]. However, there is a conflict between securing data aggregation and protecting data confidentiality, as in order to best aggregate data, it is usually necessary to process decrypted plaintext data. One framework that can be used for secure data aggregation is TAG, which uses an SQL-like language for querying clusters in the WSN [4].
For plaintext data aggregation, a secure information aggregation (SIA) framework exists [4]. In SIA clusters, there are resource-enhanced cluster heads as well as sensor nodes. Here, the cluster heads have a separate encryption key for talking to the nodes and forwarding data, for instance to a base station. Again, a combination of routing protocol and encryption mechanisms should be used for securing data processing in the network.
Known methods for aggregating plaintext data in sensor networks are either securing in-network processing (SINP) [18] , or energy-efficient secure pattern-based data aggregation (ESPDA) [19], [21], or secure differential data aggregation (SDDA) scheme based on pattern codes [20] , and lastly, witness-based data aggregation (WDA) [22].
A second approach is to provide data aggregation and process calculations on encrypted data, by using special transformation algorithms on the encrypted data. The drawback to this approach is that the security level might be reduced [4]. This is also known as a privacy homomorphism (PH). An PH implementation called concealed data aggregation (CDA) is capable of calculating the sum's and average's in hierarchical WSN's [23], or alternatively, the HSC (homomorphic stream cipher) approach can be used [24] .

1.2.2.9 Trust Management

Trust and reputation-based security frameworks can offer further protection, when even cryptographic security mechanisms fail to deliver optimal results [4]. For instance, they make it possible to judge the quality and reliability of the data provided by particular sensor nodes, as well as the health and correctness from aggregator nodes, and the network transmission quality [4]. The main drawback of such models is that they also require a certain computational overhead. Models, such as the personalized trust model "PET" [26], have been proposed by various researchers, and our detailed use cases in the next section will also feature a trust-based approach.

1.3 Security Policy Reconfiguration Solutions in WSNs

As was demonstrated in the previous section, the optimal security policy choice depends on the situation, and the situation is not always known in advance. In this section, we address this dilemma by presenting three practical solutions for dynamic reconfiguration of network security policies. These solutions have all been selected, because they take into consideration the special requirements of sensor networks, such as optimized energy consumption and low processing overhead while still providing adequate security. Therefore the examples are particularly suited for application in dynamic WSNs.

1.3.1 Effective Key Management in Dynamic Sensor Networks

Because symmetric key encryption has several drawbacks, such as having a high communication overhead, using a large memory space, not being scalable, and not being resilient to node compromise, the security researchers chose an asymmetric key encryption method [7]. The scientists chose to use an asymmetric, certificate-less public key crypthographic method (CL-EKM) for securing the dynamic WSN. The node's private key consists of a concatenation between a partial private key generated by a key generation center (KGC) and the node's own secret value [7]. CL-EKM is built by utilizing a pairing-free certificateless hybrid signcryption scheme (CL-HSC).
The network topology consists of 3 node types and 4 key types. For the nodes, there is a base station (BS), which manages the network, collects the sensor data and hosts the KGC [7]. Additionally, there are clusters managed by nodes with high processing capabilities (H-Sensors), and sensor nodes with low processing capabilities (L-Sensors). As for the key types, the first key which is established is a certificateless public/private key pair, then there is an individual node key, a pairwise key for node-to-node communication in the neighborhood, and a cluster key for clustered nodes [7].
The network itself knows seven phases [7]. In the deployment phase (phase 1), the BS sets up public/private keys for each node, creates and curates both a node list, and a node revocation list, and distributes individual keys for each node. In the pairwise key generation phase (phase 2), nodes advertise their presence to their neighbors and setup pairwise keys for communication with each other by exchanging their identifier and public key, which is verified by the BS. The nodes set up long-lived and short-lived pairwise keys with their neighbors. In the cluster formation phase (phase 3), L-Sensors cluster around H-Sensors and common cluster keys are deployed, whilst the clusters are managed by the BS. In phase 4, nodes frequently update their keys in order to protect against cryptanalysis and mitigate damage from compromised keys. Additionally, nodes can move between clusters, which might trigger cluster key renewal (phase 5). If a compromised node is detected, the BS can revoke their key's (phase 6). And in phase 7, new nodes can be introduced to the network. Once the system is setup (phases 1-3), the network is stabilized and individual nodes can fluctuate between phases 4-7.
CL-EKM was implemented with Contiki OS and TinyECC, and the performance was evaluated by using the Contiki simulator COOJA [7]. Figures 1.2 and 1.3 show the performance and power consumption of various ECC based key management schemes. This illustrates that there is a tradeoff between overall energy consumption of the node and the time a node will require to negotiate the encryption. Depending on the application needs, both algorithm, node CPU clock rate have to be selected carefully in order to achieve optimal results in terms of power consumption and time spent. Lastly, we can see in Figures 1.4 and 1.5, that once master encryption keys have been established, encryption key establishment is much less computationally intensive - demonstrating how important it is that the correct encryption methods are selected for any given communication channel.

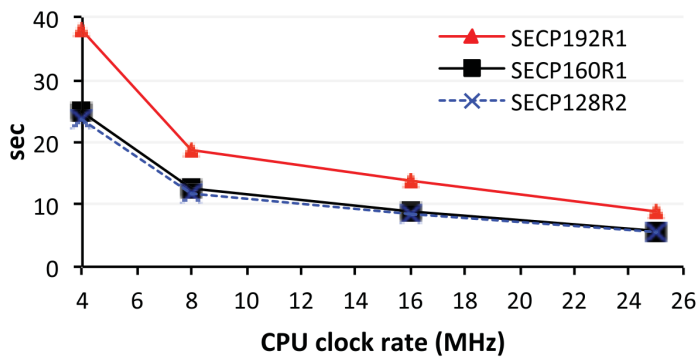


Figure 1.2: Computation overhead for pairwise master/encryption key establishment, encapsulating key information. [7]

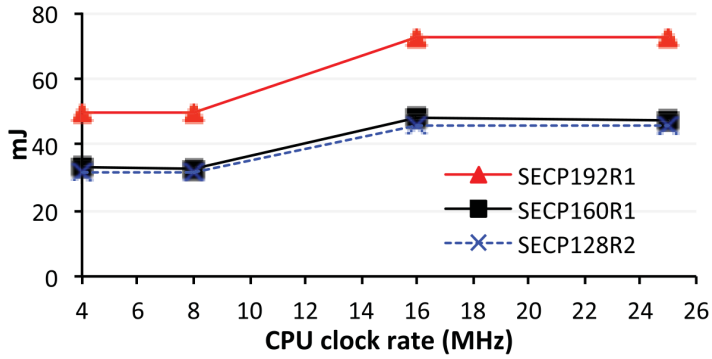


Figure 1.3: Energy consumption for pairwise master/encryption key establishment, encapsulating key information. [7]

Figures 1.6, 1.7, and 1.8 show the network power consumption for variable security configurations based on a random walk model. Figures 1.9, 1.10, and 1.11 show the same for a Manhattan mobility model. The simulations are based on a 400 x 400 m grid, with 100 L-Sensors and 25 H-Sensors deployed. The configurable security settings are a) *tbackoff* (how many seconds from a node leave until cluster heads update the cluster key, and b) *thold* (how many seconds the pairwise key is kept in memory).
The larger tbackoff, the lower the security level is. Thus, there is a tradeoff between the security level and the energy consumption of the H -sensor. However, at high speeds (>16 m/s), tbackoff should be less than 1s in order to minimize the number of cluster key updates. At low speeds, 1, 2 or 3 seconds are a reasonable choice for the H-sensors [7].
The effect of thold increases as the node speed increases. As thold increases, the energy consumption decreases because in the event that the L-sensors return to the old clusters before the timers expire, no new pairwise master key establishment is necessary [7].
Overall, this example scenario implementing a certificateless key management scheme for WSNs demonstrates how it is required to make tradeoffs between security and performance on a high level, whenever a WSN is built. The main question therefore always has to be, how sensitive the data aggregated and forwarded by the network is, and what protection efforts are necessary in order to protect that data.

1.3.2 Trust-based Enforcement of Security Policies

Our second solution aims to optimize resource consumption for security policies. The *Trust Based Enforcement of Security Policies*, or TESP, is designed and presented by Robert Vigo, Alessandro Celestini, Francesco Tiezzi, Rocco de Nicola, Flemming Nielson, and Hanne Riis Nielson[8]. This group aims to dynamically determine the perfect balance

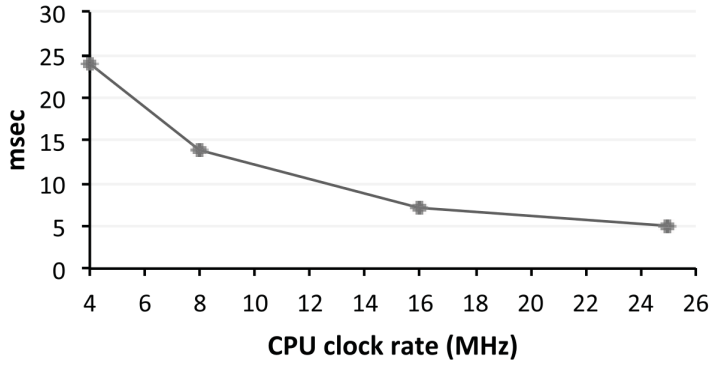


Figure 1.4: Computation overhead for pairwise master/encryption key establishment, pairwise encryption key establishment. [7]

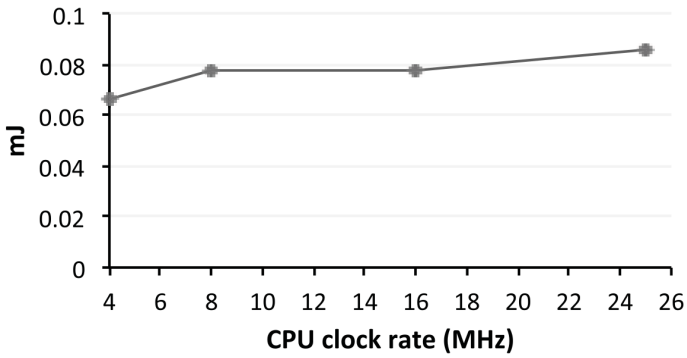


Figure 1.5: Energy consumption for pairwise master/encryption key establishment, pairwise encryption key establishment [7]

between resource consumption and security strength through a system of reputations. Put in their own words, they *"propose a modelling language and a framework in which security checks can be relaxed or strengthened to save resources or increase protection, on the basis of trust relationships among communicating parties. Such relationships are automatically derived through a reputation system, hence adapt dynamically to the observed behavior of the parties and are not fixed a priori"* [8].

The authors define security polices as *"fixed rules concerning actions that have to be executed whenever given conditions are met"*. The situations in which the security policies are utilized are often highly dynamic and decisions may depend on factors unknown during the design. This is the case when the adversaries are humans or when the system in question is highly limited in resources - for wireless latter networks, the latter is an inherent issue. The authors address the dynamic nature of threats in networks by allowing the probability of strengthening a policy to change as we gain information: *"the probabilistic enforcement of security policies can account for relaxing or strengthening a security check, aiming at saving re- sources or increasing protection. In order to react to a dynamic environment, the probability to undertake, relax, or strengthen a given security measure cannot be fixed a priori, [...] but depends on our knowledge of the past and on our expectations for the future"* [8].

The system uses a probabilistic reputation systems, where trust is quantified in probabilistic terms. Communicating parties rate each other after every interaction and then use rates to calculate a reputation score, which is used to manage future interactions. The authors add that the main advantage of reputation systems is that *"they permit computing the probability that a policy is enforced as the system evolves and according to its observed behaviour. This leads to the notion of trust-based enforcement of security policies. The*

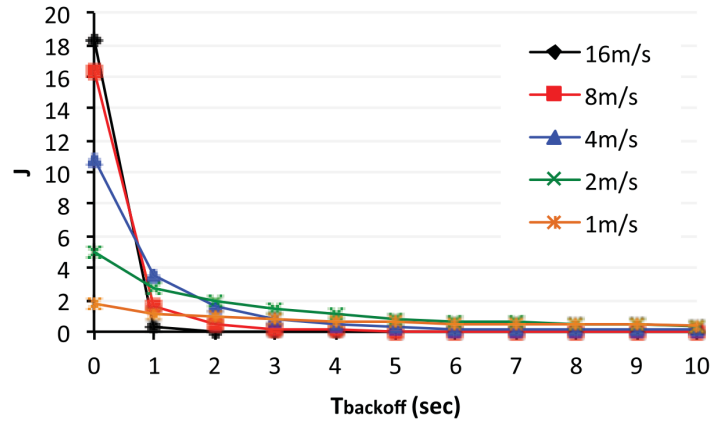


Figure 1.6: Node movement simulation results in random walk mobility model. Energy consumption of one H-sensor for cluster key update for one day (Thold = 100 sec). [7]

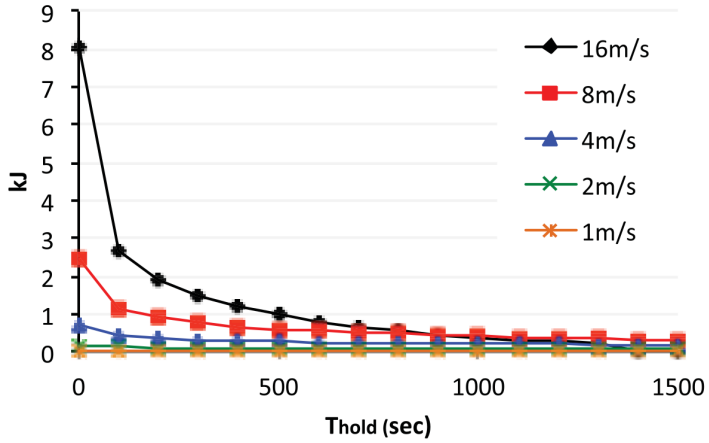


Figure 1.7: Node movement simulation results in random walk mobility model. Energy consumption of one H-sensor for pairwise key establishment for one day (Tbackoff = 6 sec). [7]

trust we put on the involved parties at a given time can be used to determine the actual enforcement of security policies" [8].

With this system in place, policies can be relaxed to save resources: expensive checks can be skipped or relaxed with interacting with a party that has shown fair behavior in the past. On the contrary, we can strengthen a policy when interacting a party we don't trust, or we can limit exchange of critical information to highly-trusted parties.

The group's solution to the challenge of modelling such a system is TESP (Trust-based Enforcement of Security Policies), a *"calculus inspired by [DTU Compute, Technical University of Denmark], that enriches the StoKlaim formalism with probabilistic aspects for security policies and with primitives for managing reputation scores. The main technical challenge addressed in this work is the integration of all these elements (in particular, probabilistic aspects, security policies, and reputations) in a uniform framework. TESP (Trust-based Enforcement of Security Policies) is a distributed process calculus which can be used to model security policies and their probabilistic enforcement, relying on a reputation system to infer probabilities. The calculus, displayed in Figure 1.12, enriches (a subset of) StoKlaim with primitives for handling reputations and inherits the aspect-oriented mind-set of AspectKP for the design of security policies"* [8].

The base station is in charge of receiving readings and sending them to the central server. Once the signature check is successfully passed, the sensor reading is forwarded. As described by the authors, the process *"consumes a check request, compares the signature*

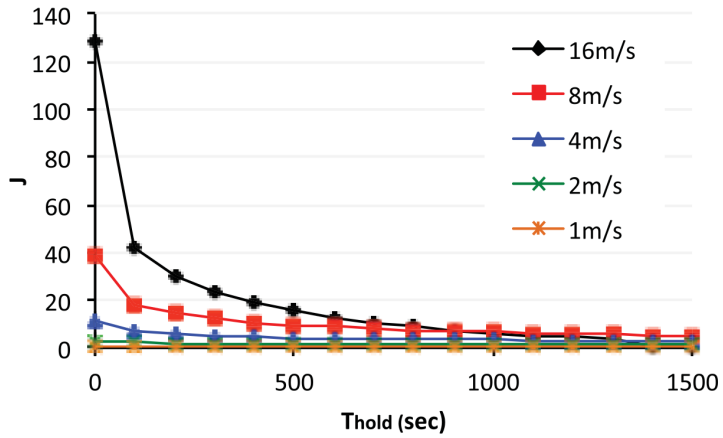


Figure 1.8: Node movement simulation results in random walk mobility model. Energy consumption of one L-sensor for pairwise key establishment for one day (Tbackoff = 6 sec).[7]

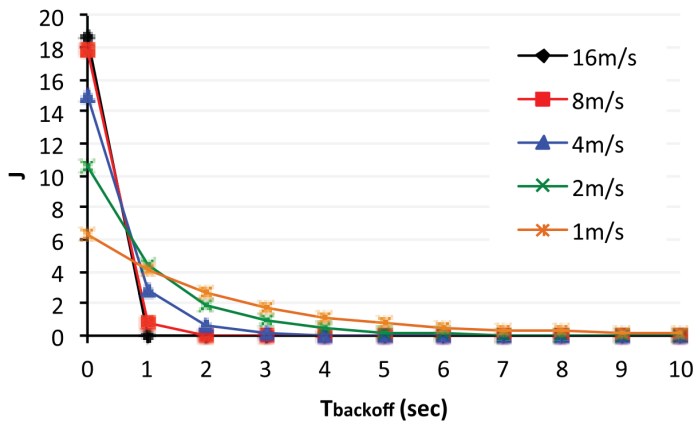


Figure 1.9: Node movement simulation results in Manhattan mobility model. Energy consumption of one H-sensor for cluster key update for one day (Thold = 100 sec). [7]

with the source locality and sends a Boolean value representing the result of the comparison to the requester. In case the check fails, i.e., the message is corrupted, the base station discards it, otherwise it is forwarded to the server." The sensor receives a positive score if the check succeeds and a negative score if it fails. The reputation score is set to the number of positive interactions divided by the total number of interactions. This naturally raises the question - what is the score of each node in the beginning? To be safe, the score of all nodes is set to 0, where 0 is complete distrust and 1 is complete trust.
With a reputation system in place, the base station chooses for each new interaction whether no policy is enforced (action denied at source), one policy is enforced (action allowed at source and no policy at target), or two policies are enforced (both at source and target). This decision is made based upon the given threshold of interest. In the following table, we can see the probabilities of forwarding a corrupt message for different thresholds. As the authors point out, *"we observe that the probability of satisfying the formula is strictly related to the sensor's behaviour and to the threshold value. In particular, it is not always true that the better the behaviour of the sensor, the lower the probability of forwarding corrupted messages, as one might expect. For th = 1% the best result - that is, the lowest probability of sending more than 1% corrupted messages - is achieved when the sensor's behaviour is the worst (θ = 0.2). This is due do to the fact that messages from trusted parties are seldom checked, yielding a higher number of forwarded messages".* After simulating the process with SAM, two figures were ploted - the first with the ration

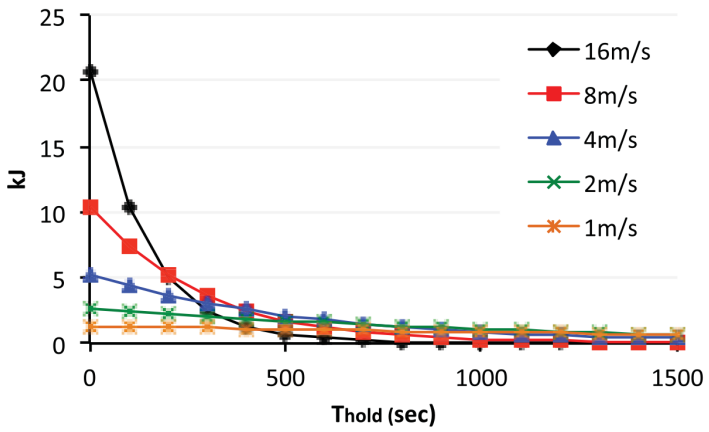


Figure 1.10: Node movement simulation results in Manhattan mobility model. Energy consumption of one H-sensor for pairwise key establishment for one day (Tbackoff = 6 sec). [7]

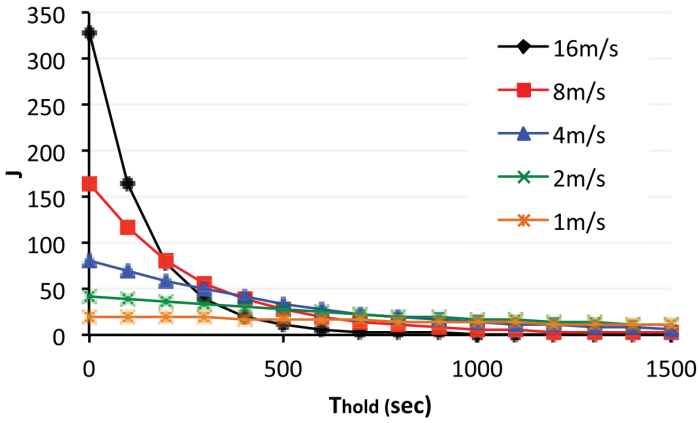


Figure 1.11: Node movement simulation results in Manhattan mobility model. Energy consumption of one L-sensor for pairwise key establishment for one day (Tbackoff = 6 sec).[7]

between the number of unchecked and forwarded messages, and the second with the number of corrupted and forwarded messages.

As we expect in Figure 1.13, we see that the better the sensor behavior, the higher the ratio of unchecked messages, and thus the more resources saved. However, Figure 1.14 seems not to confirm the results of the model checking analysis: the ratio of corrupted messages over the forwarded is lower for $\theta = 0.8$ than for $\theta = 0.2$. Nonetheless, the authors urge us to mind that *"the simulation computes an average value, while the model checking algorithm estimates the probability of exceeding a threshold. Hence, we can conclude that on average the ratio between corrupted and forwarded messages is lower for $\theta = 0.8$, but the probability that the number of corrupted messages exceeds a percentage of the forwarded messages is lower for $\theta = 0.2$"* [8]. This is due to the decreasing number of checks that are performed for $\theta = 0.8$, and to the fact that the model checking analysis has a temporal horizon: for a fixed time limit t, the total number of messages processed by the system is higher for $\theta = 0.8$ than for $\theta = 0.2$.

Through this reputation system, we dynamically adjust the probability that a security policy will be strengthened or weakened. In this way, we find an optimized balance between security strength and resource usage while adapting to the current environment and security needs of the WSNs.

$$\pi ::= \mathsf{do} \quad | \quad \mathsf{skip} \quad | \quad [?_{rep(u)} rec \ \mathsf{if} \ cut : cond] \quad | \quad \pi_1.\pi_2 \qquad \text{(Policy)}$$

$$cut ::= l :: a@u \qquad \text{(Aspect trigger)}$$

$$cond ::= \mathsf{true} \quad | \quad \mathsf{false} \quad | \quad d_1 = d_2 \quad | \quad \mathsf{readp}(T)@u : \lambda \qquad \text{(Applicability condition)}$$
$$| \quad \neg cond \quad | \quad cond_1 \wedge cond_2 \quad | \quad cond_1 \vee cond_2$$

$$rec ::= \mathsf{skip} \quad | \quad \mathsf{do} \quad | \quad P \qquad \text{(Recommendation)}$$

Figure 1.12: Syntax: Aspects and Policies [8]

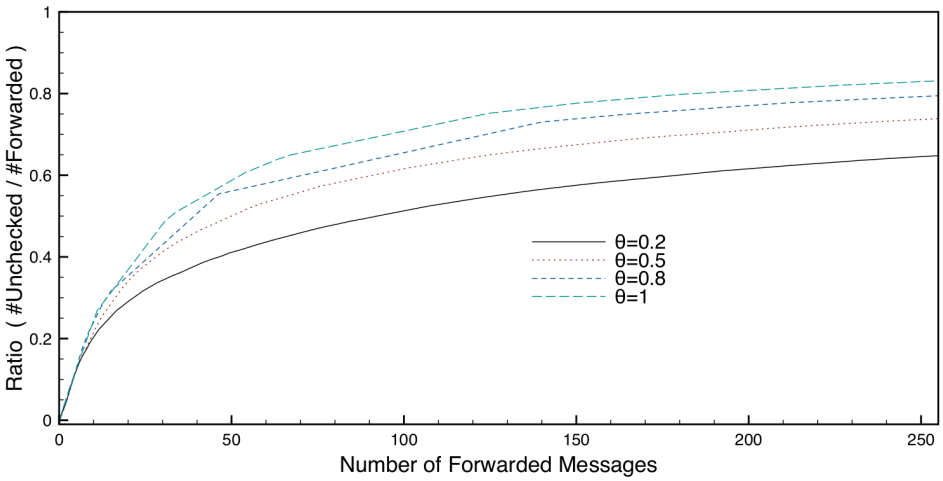


Figure 1.13: Ratio between unchecked and forwarded messages [8]

1.3.3 Dynamic Reconfiguration - Inter-trust & Famiware

This section introduces the third approach for the security policy reconfiguration. Here, we first discuss the dynamic software product lines (DSPLs) in FamiWare. It primarily discusses dynamic software product lines and dynamic security frameworks. We focus on how these techniques are used in FamiWare and Inter-trust, respectively. Later in this section we give the integration of Famiware and Inter-trust Framework and demonstrate why the Inter-trust framework is effective in achieving the dynamic reconfiguration of security policies.

1.3.3.1 DSPLs in FamiWare

The WSN nodes have low-level programming, limited capabilities, and different protocols for routing. This means that the system needs some self-adaptation and self-protection mechanisms, which can be achieved by using the dynamic adaptation of the algorithms used for security. DSPLs represent an emerging field that is producing software products capable of adapting to requirements that change at runtime [10]. DSPLs allow managing the software adaptation and can therefore be considered the key technologies for building self-protected WSNs with different approaches for each configuration challenge.
The dynamic approach can be applied by using the DSPLs (FamiWare Middleware) and the Inter-Trust framework in the security reconfiguration. This is to solve the existing security issues, by making each node the decision maker for adapting the security policies according to the physical environment.
A sensor node is a composition of features from asset of common and variable features, where the variable features may vary at runtime. Therefore, the products have the capacity to reconfigure themselves by binding variation points at runtime, several times,

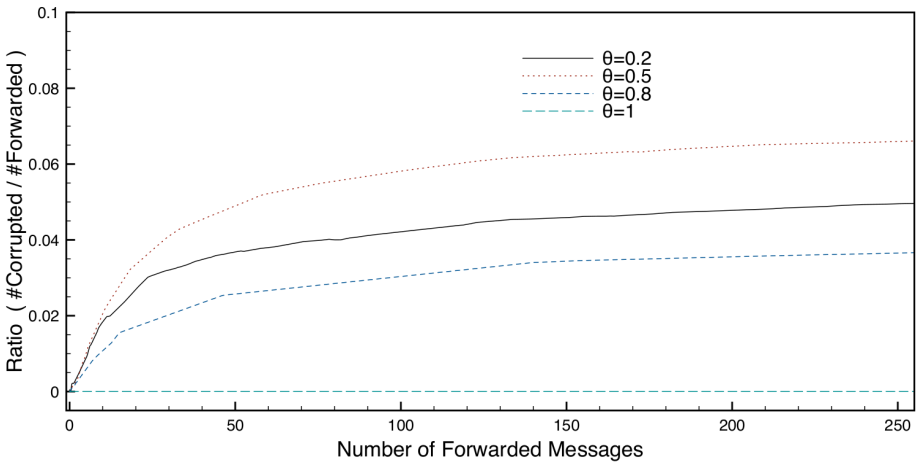


Figure 1.14: Ratio between corrupted and forwarded messages [8].

in accordance with the changes in the environment. *"DSPLs uses variability model to specify the nodes that can be changed at runtime. In the solution, the use of non-standard languages or formalisms is avoided, and we use the Common Variability Language (CVL) proposed as the standard by the Object Management Group (OMG)"*[10]. CVL is a domain-independent language for specifying and resolving variability. The CVL variability models allow modeling the variability separately from the base model, but both the variability and the base models are connected and can be managed using the same tool. The main characteristics of CVL are used to model the FamiWare variability model. The dynamic reconfiguration is defined as replacing one algorithm with other depending on the environment change. As the context changes, the nodes are bound to change to get the perfect configuration. Therefore, using the CVL approach, we automatically generate both the initial configuration and also the successive adapted configurations.
"The inherent heterogeneity and the variability of WSNs are modeled in FamiWare using CVL variability models (i.e., the VSpec trees)"[10]. FamiWare proposes the customization of the piece of software related to the middleware platform that is deployed in each device of the WSN according to:

1. the device features (device [1..*] feature in Figure 1.15 with a 1..* cardinality to represent the different devices in the system);

2. the application requirements (application feature); and

3. the global characteristics of the network used to connect the different devices (network feature)

Each device has some other features, and can be represented by its types and the FamiWare services that need to be instantiated in that particular device according to the applications requirements. Regarding their types, devices can be of high capacity computers and other sensors. For FamiWare, the variability is decided for each device, so the FamiWare services instantiated in each device can be different. Among other services, FamiWare is composed of three optional services: monitoring, context awareness, and reconfiguration. The FamiWare is checked on different types of characteristics such as memory consumption, battery level, and status of network connections.

1.3.3.2 Dynamic Security Frameworks: Inter-trust

The security framework is a reusable set of modules that can be customized to develop trustworthy applications. The framework describes the functionalities of each module that can be distinguished in a security framework. There are two different frameworks:

1. A framework that contains a set of modules which implements authorization, authentication, privacy, and other such security concerns.
2. A framework containing modules which deploys the security applications.

Among these modules, the framework can incorporate specific modules to adapt the security level of application at runtime. In this scenario we use dynamic security framework. The Inter-trust framework is a dynamic security framework [12] which supports trustworthy applications based on the enforcement and dynamic adaptation of security policies. Here, security policies are first specified using a security editor (e.g., MotOrBac, an Organization-Based Access Control policy editor) and then negotiated between the different parties in a communication using a negotiation module. When the security requirements change at runtime, the security policies are renegotiated. The negotiated security policy is then analyzed and interpreted by the policy engine and the policy interpreter modules. These modules are responsible for identifying changes in the security policy that require adapting the security concerns (e.g., authentication, encryption) deployed inside the application dynamically, at runtime.

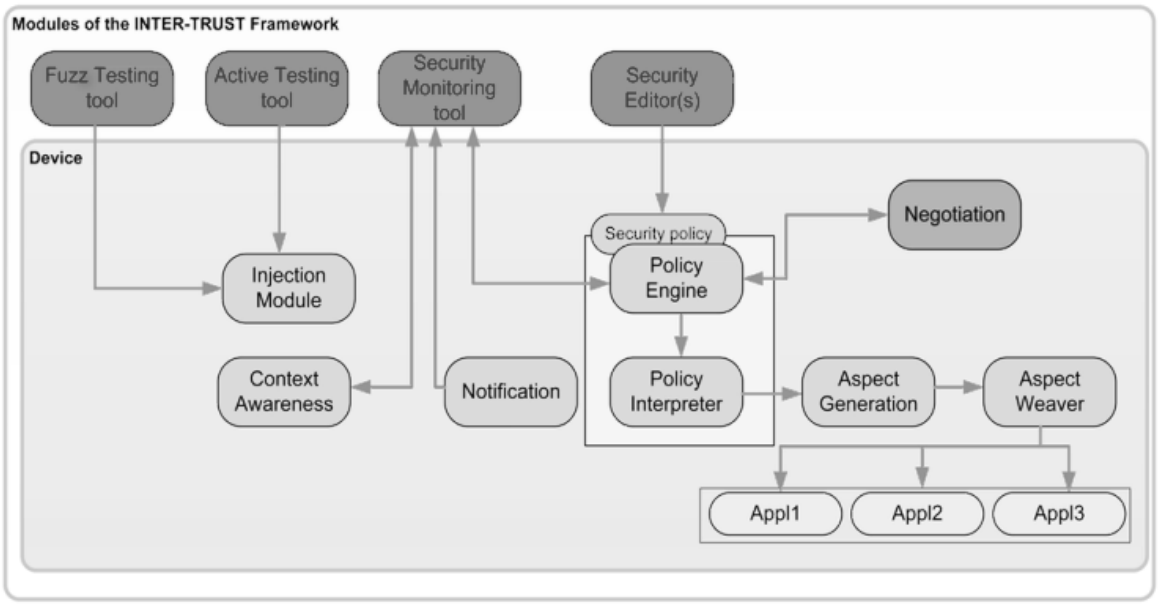


Figure 1.15: The Inter-trust framework [12].

1.3.3.3 Integration of FamiWare and Inter-trust framework

FamiWare is a family of middlewares specially designed and implemented for dealing with the heterogeneity of WSNs devices. It also provides a mechanism to be reconfigured at runtime when the context changes. However, it does not focus on providing security. Therefore, if we want to extend FamiWare with the benefits provided by the INTER-TRUST framework, the first thing to do is to integrate INTER-TRUST into the FamiWare family. There are two main steps for integration of the two application models, the first step is the integration of the INTER-TRUST modules and the INTER-TRUST security concerns into the FamiWare product line. This implies the extension of both the variability model and the base model of FamiWare.

1.3.3.4 Dynamic Reconfiguration of Security in FamiWare

FamiWare uses VSpec tree at runtime to handle the reconfiguration and the task to change the policies as the context changes. It uses models@run.time [11] approach to interpret

the reconfiguration plans with the task to complete for switching from the middleware configuration to the new configuration of policy. These changes are performed, taking care of the architectural design, and thus most of the time fits the system restrictions. Other approaches for reconfiguration include working on code level, where the issue of hard-coding arise, so the competence between the new code and the models used for design is lost. This makes it difficult to claim that the reconfiguration applied at one node will work properly for whole network, when it is deployed at each node.

1.3.4 Evaluation of Inter-trust and FamiWare

This Section talks mainly about the Challenges and Approaches for INTER-TRUST & FamiWare. Here we present the different cases and the problem areas with respect to the existing scenarios with the WSNs, and how the given approach gives the solution for it.

1.3.4.1 Challenges and Approaches for Inter-trust and FamiWare

The main challenges that need to be taken into account in order to add security to WSNs and to dynamically (re)configure the security level at runtime, can be described as follows:

1. *"Build secure applications that run on heterogeneous nodes with limited capabilities"* [10].

   As we have seen in WSNs each node is different and provides different services. At each node there are different devices and applications installed that makes the node heavier and unable to adapt the different algorithms. Here we are giving the solution to this by integrating the INTER-TRUST framework and FamiWare security product lines. The nodes should not carry other things than the algorithms to switch as the context changes.

2. *"Dynamic negotiation of security policies for WSNs.*

   The conditions of the running environment of the WSNs applications could change. Then, the security policies may need to be adapted to the new conditions. This adaptation requires a previous negotiation between the different parts of the application" [10]. The solution we present use the same integration of approaches. By this integration FamiWare middleware gives the platform for dynamic negotiation for security policies for the sensor nodes that are already included into the FamiWare.

3. "Monitor the changes in the context.

   These changes can be general (e.g., changes in the amount of memory or the battery level available) or can be specific to security. In order to react to these changes, the environment needs to be monitored" [10]. The solution can be again obtained from choosing the different essential monitoring features available in both the applications as Inter-trust and FamiWare. Here we will only choose those monitoring modules which are essential for particular node, by choosing only those important components from Inter-Trust which is not present in FamiWare.

4. *"Define a dynamic reconfiguration service to endow applications with self-protection. When the environment or the security requirements change at runtime, the system must be self-protected to maintain a proper and secure functioning"* [10].

   "As WSNs are composed of a large number of tiny devices, the self-protection mechanism should consume the least number of resources as possible" [10]. Our approach says, it is proven that the FamiWare pays special attention to the efficiency of the nodes. The resource consumption of Famiware is optimal for certain typical type of sensors, and it is also scalable to a number of nodes.

1.4 Summary and Conclusion

This paper combines a general overview of WSNs, including their applications, vulnerabilities, and theoretical defense mechanisms, with three practical solutions for dynamic security policy reconfiguration. Although research and development is providing us with stronger and more efficient security policies, utilizing these mechanisms in the highly limited-resource nodes remains a challenge. Not only are WSNs constrained by power, memory, and processing ability, but also the nodes are frequently deployed in remote areas such as forests and earthquake-prone areas. The nodes are vulnerable on four levels - the physical, link, network, and transport layers. Currently, mechanisms exist for many of the security issues. These solutions range from arming the node in tamper-proof hardware to requiring clients to solve puzzles. However, there remains the challenge of balancing limited resource and security strength. In an unknown environment where threat level and type may constantly shift, this problem is exacerbated. With this as motivation, we present three solutions for dynamic reconfiguration - an approach which aims to optimize resource usage and provide the level of security dictated by the situation at hand.

The practical examples are complete implementations for providing end-to-end security in dynamic wireless sensor networks. The first approach utilizes asymmetric certificate-less key management, implemented with Contiki OS and TinyECC. Adjusting the speed of cluster key updates tunes the security strength, allowing the user to find a fitting balance between security and performance on a high level. The calculus-based TESP mechanism, implemented in StoKlaim, proses a probabilistic framework for dynamically adjusting the security strength based on the reputation score of the nodes involved. The final implementation combines two existing mechanisms - INTER-TRUST and FamiWare. Of the three dynamic reconfiguration solutions we analysed, this is the most sophisticated approach. INTER-TRUST provides a framework for dynamic negotiation and deployment of security policies, while FamiWare manages the configuration and reconfiguration of middleware in WSNs. The fusion of these mechanisms results in a solution for the dynamic reconfiguration of security policies in a heterogeneous node network.

The implementations for dynamically reconfiguring a wireless network are quite new - the third approach was presented only a year ago. We anticipate the development of reconfiguration mechanisms with increased efficiency and plasticity. In an era where wireless technology is becoming prevalent in greater depth and scope in our every day lives, security becomes ever more crucial. With that, we would like to close with the following quote, which already hints at a future where WSN's are distributed all over the world:

"Sensor networks are set to become a truly pervasive technology that will affect our daily lives in important ways. We cannot deploy such a critical technology, however, without first addressing the security and privacy research challenges to ensure that it does not turn against those whom it is meant to benefit " [2].

Bibliography

[1] M. Waldburger, P. Poullie, B. Stiller: *Guideline for Seminar Reports*, Communication Systems Group, Department of Infromatics, University of Zurich, January 2013. http://www.csg.uzh.ch/teaching/guideline-seminar-report-v05.pdf.

[2] Chan, Haowen, and Adrian Perrig: *Security and privacy in sensor networks*, Computer 36.10 (2003): 103-105. http://www.netsec.ethz.ch/publications/papers/ieee-secure-sensor-nets.pdf.

[3] Y. Wang, G. Attebury, and B. Ramamurthy: *A survey of security issues in Wireless Sensor Networks*, IEEE Communications Surveys and Tutorials, Vol. 8, No. 2, pp. 2-23, 2006.

[4] J. Sen: *A Survey on Wireless Sensor Network Security*, International Journal of Communication Networks and Information Security (IJCNIS), vol.1, no.1, pp. 55-78, August 2009.

[5] ITU-T: *Y2060 - Overview of the Internet of things*, 2012. https://www.itu.int/rec/T-REC-Y.2060-201206-I.

[6] M. Pinto, N. Gamez, L. Fuentes, M. Amor, J.M. Horcas, I. Ayala: *Dynamic Reconfiguration of Security Policies in Wireless Sensor Networks*, Sensors, vol. 15, 2015, pp. 5251-5280. DOI: 10.3390/s150305251. http://www.mdpi.com/1424-8220/15/3/5251/htm.

[7] Seo, Seung-Hyun, J. Won, S. Sultana, and E. Bertino. *Effective key management in dynamic wireless sensor networks*, Information Forensics and Security, IEEE Transactions on 10, no. 2 (2015): 371-383. http://docs.lib.purdue.edu/cgi/viewcontent.cgi?article=1640&context=ccpubs.

[8] Vigo, Roberto, A. Celestini, F. Tiezzi, R. De Nicola, F. Nielson, and H. R. Nielson. *Trust-based Enforcement of Security Policies*, In Trustworthy Global Computing, pp. 176-191. Springer Berlin Heidelberg, 2014. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.431.4963&rep=rep1&type=pdf.

[9] Erfani, S. Hossein, H. H. S. Javadi, and A. M. Rahmani. *A dynamic key management scheme for dynamic wireless sensor networks*, Security and Communication Networks 8, no. 6 (2015): 1040-1049.

[10] M. Pinto *, N. Gámez , L. Fuentes , M. Amor , J. M. Horcas and I. Ayala. *Dynamic Reconfiguration of Security Policies in Wireless Sensor Networks.* Sensors 2013, 15(3), 5251-5280. [Google Scholar]

[11] G. Blair and N. Bencomo, Lancaster University Robert B. France, Colorado State University. *models@run.time.* 2009 IEEE, IEEE Computer Society. [IEEE xplore]

[12] Ayed, S. Telecom Bretagne, Rennes, France Idrees, M.S. ; Cuppens-Boulahia, N. ; Cuppens, F. ; Pinto, M. ; Fuentes, L.*Security Aspects: A Framework for enforcement of security policies using AOP.*Signal-Image Technology & Internet-Based Systems (SITIS), 2013 International Conference on. [IEEE xplore]

[13] M.O. Rabin, *Digitalized Signatures and Public-Key Functions as Intractable as Factorization*, Cambridge, MA, Technical Report, 1979.

[14] J. Hoffstein, J. Pipher, and J.H. Silverman, *Ntru: A ring-basedpublic key cryptosystem*, In Proceedings of the 3 rd International Symposium on Algorithmic Number Theory, London, Springer-Verlag, 1998, pp. 267-288.

[15] R.L. Rivest, A. Shamir, and L. Adleman, *A method for obtaining digital signatures and public-key cryptosystems*, Communications of the ACM, Vol. 26, No. 1, pp. 96-99, 1983.

[16] V.S. Miller, *Use of elliptic curves in cryptography*, In Lecture Notes in Computer Sciences: 218 on Advances in Cryptology-CRYPTO 85, New York, Springer-Verlag, 1986, pp. 417-426.

[17] N. Kobiltz, *Elliptic curve cryptosystems*, Mathematics of Computation, Vol. 48, pp. 203-209, 1987.

[18] J. Deng, R. Han, and S. Mishra, *Security support for in-network processing in wireless sensor networks*, In Proceedings of the 1st ACM Workshop on Security of Ad hoc and Sensor Networks, New York, ACM Press, 2003, pp. 83-93.

[19] H. Cam, D. Muthuavinashiappan, and P. Nair, *ESPDA: Energy-efficient and secure pattern-based data aggregation for wireless sensor networks*, In Proceedings of IEEE Sensors, Toronto, Canada, October 2003, pp. 732-736.

[20] H. Cam, S. Ozdemir, H.O. Sanli, and P. Nair, *Sensor Network Operations*, Wiley, 2004, Chapter: Secure Differential Data Aggregation for Wireless Sensor Networks.

[21] H. Cam, D. Muthuavinashiappan, and P. Nair, *Energy-efficient security protocol for wireless sensor networks*, In Proceedings of IEEE VTC Conference, Orlando, Florida, October 2005, pp. 2981- 2984.

[22] W. Du, J. Deng, Y.S. Han, and P.K. Varshney, *A witness-based approach for data fusion assurance in wireless sensor networks*, In Proceedings of IEEE Global Telecommunications Conference, San Francisco, December 2003, pp. 1435-1439

[23] J. Girao, D. Westhoff, and M. Schneider, *CDA: Concealed data aggregation for reverse multicast traffic in wireless sensor networks*, In Proceedings of IEEE International Conference on Communications, Seoul, Korea, May 2005.

[24] C. Castelluccia, E. Mykletun, and G. Tsudik, *Efficient aggregation of encrypted data in wireless sensor network*, In Proceedings of ACM/IEEE Mobiquitous, San Diego, July 2005.

[25] A. Seshadri, A. Perrig, L. Van Doorn, and P. Khosla,*SWATT: Software-based attestation for embedded devices*, In Proceedings of the IEEE Symposium on Security and Privacy, May 2004.

[26] Z. Liang and W. Shi, *PET: A Personalized Trust model with reputation and risk evaluation for P2P resource sharing*, In Proceedings of the HICSS-38, Hilton Waikoloa Village Big Island, Hawaii, January 2005.

Chapter 2

Private Cloud: Technology and Deployment

Geetha Parangi, Linda Diep, Elfat Esati

Cloud Computing is a primordial change that provides computing over the Internet to meet upcoming demand of computing and storage resources for organisation without huge investment. There are several reasons to build private clouds for security, privacy and data storage management. In this paper we propose an additional model called Home cloud as a special case of private cloud. We focus on home cloud setup using Open Source Software such as Openstack and Virtualization to access digital assets and services via apps across multiple devices in a seamless manner. We will also discuss technical strength and weakness of Openstack cloud software and Virtualization for implementing home cloud infrastructure. We analyze security threats and propose security levels and security framework and related advantages and disadvantages of setting up the home cloud.

Contents

2.1 Introduction **33**
2.2 Cloud Computing **33**
2.2.1 Challenges and Advantages of Cloud Computing 33
2.2.2 Public Cloud 34
2.2.3 Private Cloud 35
2.2.4 Hybrid Cloud 35
2.2.5 Community Cloud 35
2.2.6 Storage Versus Cloud Computing 35
2.3 Private Cloud **36**
2.4 Home Cloud **37**
2.5 Home Cloud with OpenStack **37**
2.5.1 Virtualization and Cloud Computing 37
2.5.2 Hypervisors 37
2.5.3 Architecture of OpenStack 38
2.5.4 Community App Catalog 41
2.5.5 Limitations of OpenStack 41
2.6 Reachability of Home Cloud **42**
2.6.1 Static IP Address 42
2.6.2 Static IP Address and Domain Name 43
2.6.3 Dynamic IP Address and Domain Name 43
2.7 Cost Estimation for Home Cloud Setup **44**
2.8 Security **45**
2.9 Security Risk **46**
2.10 Advantages and Disadvantages of Home Cloud **48**
2.11 Evaluation and Conclusion **48**

2.1 Introduction

Cloud computing has gained a lot of popularity among both companies and individual users for accessing computing resources via the Internet. It is a technology that has become more essential to businesses and private persons.
There are four deployment models which are public clouds, private clouds, hybrid clouds and community clouds. However in our report, we are covering the private cloud for the home use, therefore we define it as a ”home cloud”. Also, the main focus of this report shall be given to home clouds, which we see as a special case of a private cloud rather than an additional deployment model to the four above.
There are many options of how to setup a home cloud, but in this report we will zoom in on the Open Source cloud computing software OpenStack in our setup. Benefits and drawbacks of using OpenStack will then be illustrated. We also provide three possibilities of how the home cloud could be reached outside of the home network. A rough cost estimation for home cloud setup will be also given.
Cloud computing does not only come with benefits and convenience, but also with some major drawbacks. End-users have to entrust their data not only to the cloud service providers, but often also to third parties who own the physical cloud infrastructure [1]. Thus security and privacy issues arise naturally. As security is one of the main arguments why people should use home clouds instead of public clouds, we will discuss some of the security issues and challenges that come along with cloud computing and also security risks in public clouds in comparison with home clouds.
The authors looked for common and different characteristics of public and home clouds. Also special attention was paid to aspects of public cloud services and if they are also applicable on home clouds. At the end of this report, an overview of advantages and disadvantages of home clouds are given.

2.2 Cloud Computing

The concept of cloud computing did not arise recently, but it is has been emerged from many previous ideas and concepts like virtualization, distributed computing, web and software services [15].The use of this term is not always accurate, as almost any kind of web services are called a ’cloud’ nowadays. After in depth research, it seems to the authors that ’cloud’ is often misused as a buzz word in several cases to generate marketing hypes. Even though there is no standard definition for the term ’cloud computing’, the most common one is given by the National Institute of Standards and Technology [5]:
”Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers storage, applications, and services) that can be rapidly provisioned an released with minimal management effort or service provider interaction. This cloud model is composed of five essential characteristics, three service models and four deployment models **??**.” An overview of the key components are shown in Figure 1.1.

2.2.1 Challenges and Advantages of Cloud Computing

Cloud computing has changed the way IT services and resources are delivered to the customers, with the advantage of high scalability, flexibility, reliability and availability and no advance cost in acquiring and managing IT infrastructure it is adopted by enterprise, small business and also by the individuals to set up the home cloud. Some of the providers of cloud services are Amazon, Salesforce, Google and Microsoft. According to International Data Corporation for the year 2015 there will be 35 percent increase in

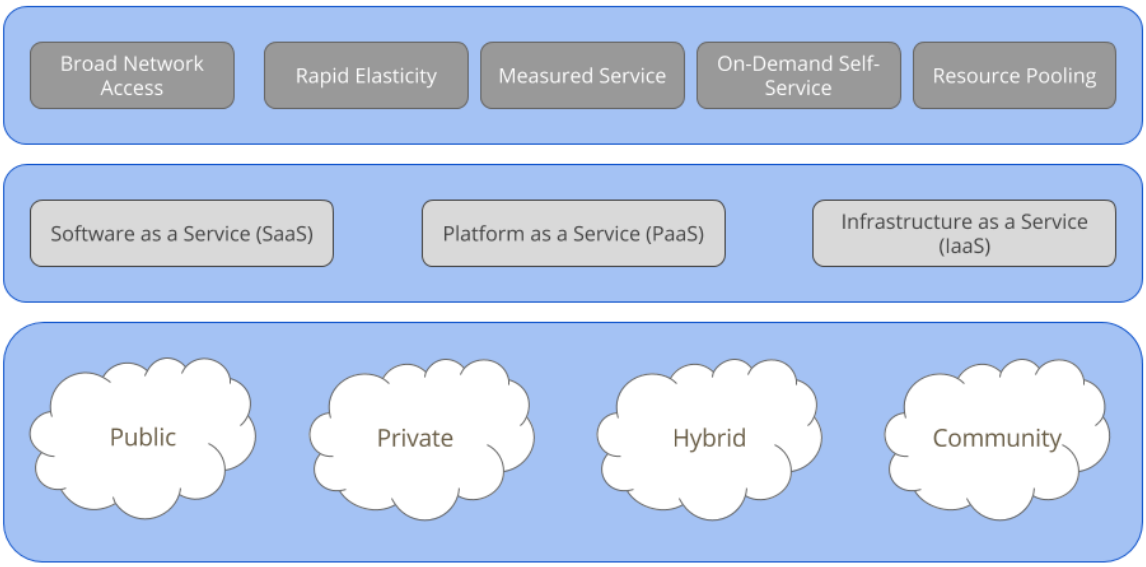


Figure 2.1: The five essential characteristics, three service models and four deployment models in cloud computing [32].

cloud services and technologies that enables these services and is expected to cross 150 billion dollars [12].
Some of the added advantages of cloud computing are:

- The organization can use how much and when they want according to current needs and scale.
- With less investment the organization can focus on core work by adding values to clients.
- It is more reliable to access data from anywhere over an Internet.
- It provides infinite storage capacity can also schedule backups to their data.

But the main challenges in cloud computing would be the user must be able to sustain failure of network or web service. Due to the absence of cloud computing API its difficult for the user to move from one service provider to another, and also the user might not know the physical location of data where they are stored although they have logical control over data but they lack physical control over it. Adequate security mechanism should be employed by the cloud service provider to cater the confidentiality need of customers.

2.2.2 Public Cloud

In public clouds the resources are usually offered over an Internet connection as a service. These services are accessible to the general public and users only pay for the services they actually use. An advantage of using clouds is that users do not need to buy hardware to use the services. Moreover, they can scale the use on their demands accordingly. Public cloud providers manage the physical infrastructure and pool the resources into the capacity required by their customers, therefore all customers share the same infrastructure pool with limited configuration, security protections and availability [2]. Another significant feature of public cloud is that they are mostly hosted away from customer premises and are owned by a third party organization that offers the cloud service.

2.2.3 Private Cloud

Private clouds are hosted on-premise and can only be accessed by the members of the hosting organization, specific group of people or granted third parties. The main purpose of the private cloud is not to offer a cloud service to the general public but rather to those within the hosting organization or restricted group of people [4]. This means that the computing architecture is dedicated to that one organization and is not shared with others. Private clouds are usually very costly, but is also considered more secure than the public clouds [4]. As private cloud or rather home cloud is the main model in our report, this model will be covered in more details in 2.3.

2.2.4 Hybrid Cloud

The hybrid cloud is a combination of public and private cloud therefore it is more complex than the other deployment models [3]. Its infrastructure is composed of two or more clouds where at least one is a private cloud and the other a public cloud. These two entities are bound together by standardize or proprietary technology, which allow data and application portability [3]. Usually the more critical and secure applications are run in private clouds and the less critical in the public cloud. Also the more sensitive data is stored in the private cloud, whereas the less sensitive data is stored in the public cloud.

2.2.5 Community Cloud

The community cloud is similar to a private cloud, but it is shared between two or more organizations rather than a single organization. Thus organizations that are likely to have common requirements for their cloud computing infrastructure, can share the same cloud [3]. Consequently, the organizations can agree on implementing a common infrastructure that is reliable and cost effective [6]. An example of sharing a cloud between members of a community are government authorities that share the the same cloud, where non government agencies have no granted access [4]. The data center for this deployment model can be on or off premise for security and performance reasons [3].
Although this model is still at its initial state, it has been gradually gaining popularity among start-ups and small and medium term businesses.

2.2.6 Storage Versus Cloud Computing

The term ”Cloud Computing” and ”Storage” may sound like interchangeable words referring to the same infrastructure, but the two system have less in common than the fact that they both store data.
If we take Network Attached storage as one of the storage devices available in the market it includes processor, memory, and space for hard drive storage that is connected to a local network so that it can be accessed remotely either through connected computers, wireless devices. It gives the security of physically owning the drive on which data is stored and convenience of accessing the data from anywhere. NAS devices are perfect for consumers with multiple computers if they want to stream files.
But the cloud computing is much beyond storage, even though storage devices are becoming more user friendly requiring minimal setup and configuration it still cannot compare to the ease of cloud computing service, so here no need of multiple computers as in the case of NAS, but the desktop app will offer the remote storage. As in the case of storage devices it will protect the user data against the failure of one of the drives but will not protect against the loss or damage of the storage device itself, but as in the case of cloud computing it includes backups as part of the services and also the cloud company will

ensure that the data will never be lost if there is a problem at their end. In order to buy the NAS storage devices it is expensive but as we compare with the cloud computing software it offer few gigabytes of storage for free but for the large amount of storage also the price hardly compares with the storage devices.

2.3 Private Cloud

Over the last several years, cloud service breaches have dominated and it negatively impacted their customer reputations and company's overall profit. Private cloud companies can ensure control over their own data, so therefore many companies are migrating their data centers to private cloud to run core business apps.
Private cloud is achieved within the corporate firewall under the control of the IT department, it provides more control over the company's data and ensures security. The organization implementing the private cloud is accountable for running and managing IT resources instead of passing that leadership on to a third party cloud provider.
Private cloud will be able to deliver server virtualization, benefits of automated provision of resources and the use of hardware within the IT infrastructure [9].
But at the same time it has got few pitfalls such as planning and preparation for the deployment is essential, and also the initial investment in hardware and software and security measures. The cloud also needs to be maintained and monitored for security and performance [8].
In order to launch private cloud the enterprise needs to analyse the need for private cloud, developing a plan to initiate cloud, generating cloud policies for access and security, deploying and testing the infrastructure and cloud enabling software such as "Openstack", "Cloudstack" and training employees and partners for the cloud computing project [10].

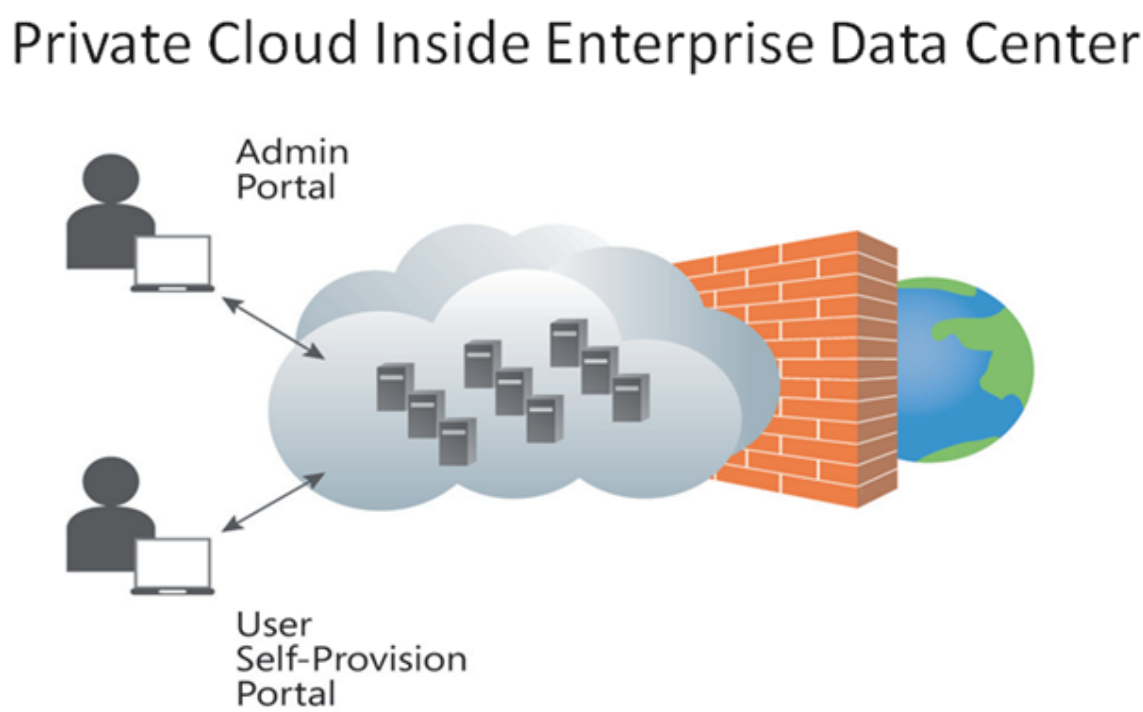


Figure 2.2: Private Cloud [33].

Private cloud is on premise deployment of cloud including the servers, cabling, software, etc that are owned and managed by the enterprise behind its firewalls as shown in Figure 2.2. To persue the cloud IT organization within the enterprise should have Iaas in first instance to provides the mechanism for resource pooling, scalability and elasticity.

In general its an expensive proposition on both cost and customer satisfaction, so the company launching the private cloud should be clear on short term checkpoints and long term business goals [11].

2.4 Home Cloud

Home cloud is a special case of private cloud but on a smaller scale such as small business or home users can set up the home cloud for sharing, storing and synchronising.
The benefits of home cloud for small, midsize business and for individual users are numerous. Help in saving business time, and money by raising productivity, improving collaboration and promoting innovation. It is a good alternative for individuals and for small business that combine the security of personal storage, synchronization, web based apps with the convenience of the cloud. It facilitates all types of information sharing such as email services, application hosting, web based phone system and much more.
Home cloud setup enable the user to access information from anywhere using any compatible device such as laptop, phone, tablet. It stores data on the Internet that will guarantee the information will be available from a central web-based hub to users with proper credentials access from any location with an Internet connection. It also syncs data for all devices connected to the cloud, keeping them updated with real-time information.
In a market there are number of open source cloud storage software that the users can use to access and sync data privately for security and privacy reasons. By setting up the home cloud there are number of advantages such as full control over the data, privacy protection, encryption, can verify the source code for bugs and backdoors, good performance as data will be stored in local storage instead of remote data center, can share the files and data with or without password or time limit, no third party corporation will own the data.

2.5 Home Cloud with OpenStack

2.5.1 Virtualization and Cloud Computing

We know that the word Cloud Computing has become quite a buzzword in the recent times. So many people might be scratching their head and tend to think of it as some kind of intangible asset that does some computing work on the fly. Basically the term covers many different technologies, among which Virtualization plays the most important role in Cloud Computing. Virtualization and Cloud Computing are used interchangeably but they do not really mean the same thing.
Virtualization is part of Cloud Computing that helps to enhance private cloud by managing the computer resources more effectively. Through virtualization we can scale up and down the computer resources with different workload, if workload were to increase without virtualization it would not be possible. In the old days the virtualization first came up when we used Terminal Server to compute all the job. As mentioned in the previous sections there are three main service models offered by Cloud Computing, but this paper focuses on Home Cloud so we will only need Infrastructure as a Service(IaaS) because it is meant for File Sharing and Synching. Based on the services the virtualization technologies differ from each other.

2.5.2 Hypervisors

To make virtualization available there is a need for Virtual Machine or Hypervisor. This software makes it possible to run multiple Operating Systems in one physical machine, as illustrated in Figure 2.3. So we can have more Operating Systems to do a particular job.

Cloud Computing infrastructure consist of multiple nodes that if we do not utilize a hypervisor than we have to install each of those nodes on a separate physical machine which is not cost-effective and we do not want that. There are different types of virtualization but in case of Home Cloud Computing with OpenStack we will focus on para-virtualization. This type of virtualization makes it possible to run OpenStack and a Hypervisor on top of the physical machine. So in this case both systems have direct access to hardware kernels. Since the host Operating System and Hypervisor communicate directly it gives a high performance without incurring any overhead due to hardware emulation [23] . In this case the guest machine is aware that it is being virtualized so it does not send commands to the hardware but instead sends command to the host Operating System.
One of the big advantages of Para-virtualization is that the Hypervisor needs little power to manage the guest machines since every guest OS is aware of each other and they know how much demand they are requesting from the host machine [24] . These were the main reasons to choose this type of Virtualization and the other types do not offer this, they simply run on top of the Host Operating System as a standalone OS that is not aware it is virtualized. But one may ask why would we need virtualization for Cloud Computing? The main reason is that OpenStack is scalable and allows to flexibly share nodes, that is why it is not meant to work with only two or three nodes.

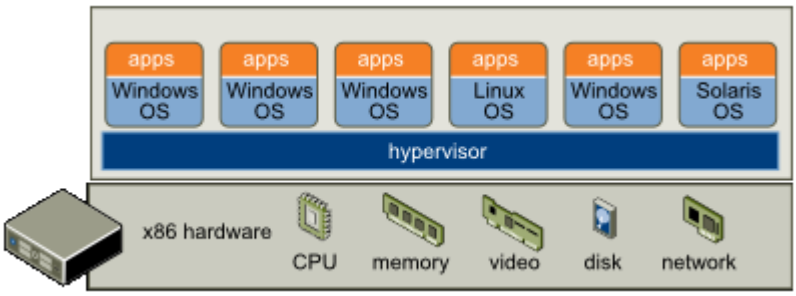


Figure 2.3: A global picture of virtualization [17]

The most popular Hypervisors that offer this type of virtualization are Xen and KVM (Kernel Based Virtual Machine). Since OpenStack is full open source the best combination would be with KVM as an open source Hypervisor as well. KVM runs almost all the Operating Systems as guests and supports x86 processors [25]. VirtualBox and Vmware are the most common Hypervisors that support full virtualization. Choosing which Hypervisor to use is quite a difficult task because the cloud performance is directly affected by it.

2.5.3 Architecture of OpenStack

In order for us to setup Home Cloud for File Sharing such as Documents, Photos, Videos and synching of them for personal usage we have chosen OpenStack as the most leading cloud operating system in use today. The OpenStack project lays its ground from NASA and Rackspace because of scalability issues NASA had to abandon their Nebula Project. There are other cloud software that provide cloud services but OpenStack provides Infrastructure as a Service (IaaS). In the following sections you will find out what one ought to know about OpenStack if we need to setup Home Cloud. Before setting up Home Cloud one ought to understand the components that make OpenStack work.

1. OpenStack Architecture

OpenStack is an open Source Cloud Operating System that delivers multiple services for enhancing small and large scale environment and separates guest Operating Systems from

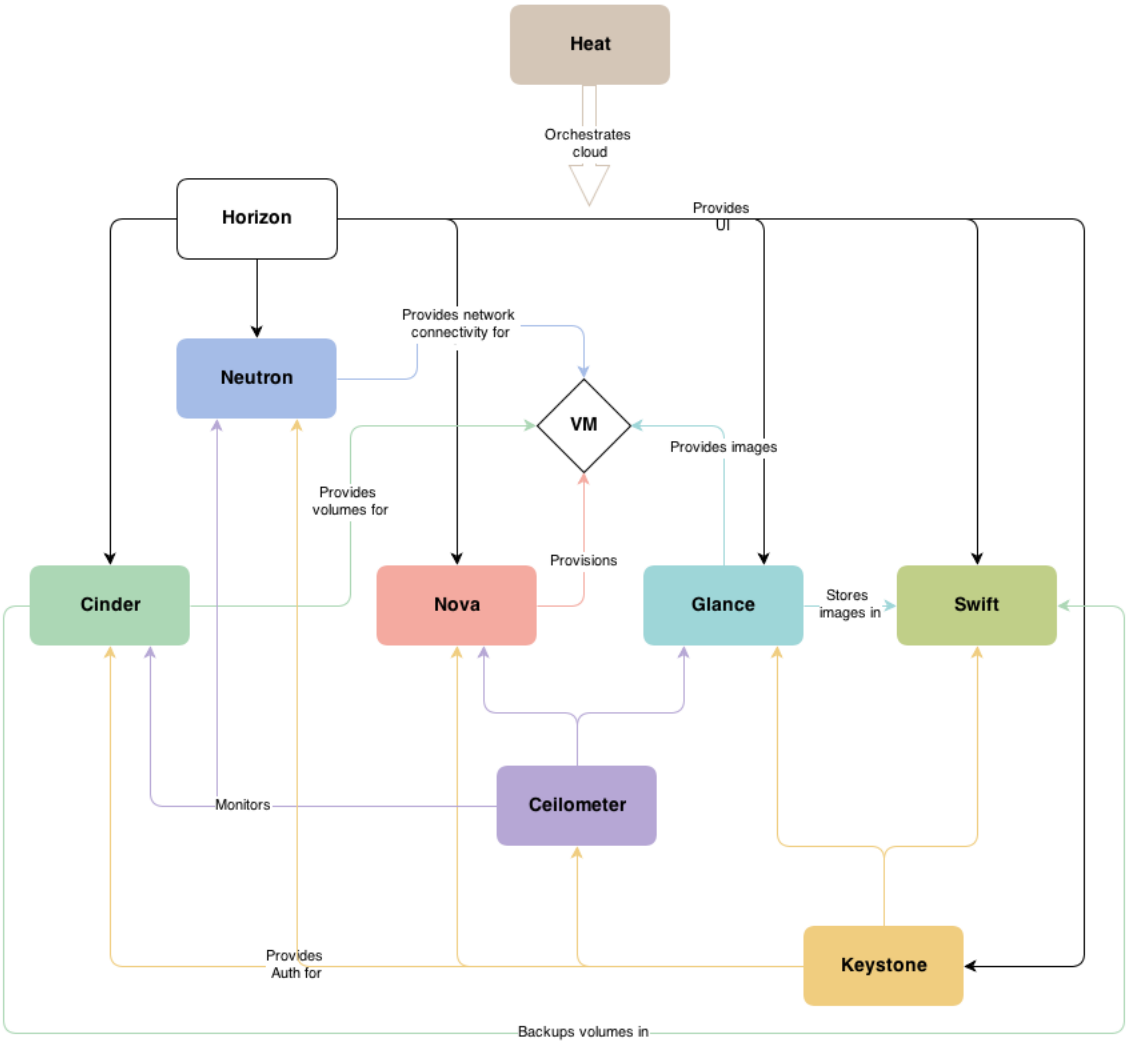


Figure 2.4: Architecture of OpenStack [18]

Applications. There are many ways to split OpenStack for deployment. But in this section we will give detailed architecture of full services as seen in the Figure 2.4.
Some say OpenStack consists of six nodes, and some say of seven nodes. The answer to this is that concurrently OpenStack has introduced new components, but in the early days there used to be only two nodes and those basic nodes were to handle the job of the current seven nodes, the two main nodes are: Controller Node, Computer Node. Concurrently we have more nodes available, but depending on the services they offer we can split them in three categories:

1. Computer Service known as Nova
2. Network Service known as Neutron
3. User Interface Dashboard known as Horizon

Compute Service Node is the core component of OpenStack which makes Cloud Computing scalable. It operates the lifecycle of the instances. Nova does a lot of services to respond to the user requests sent by the instances, furthermore Nova offers these services: Nova APIs which accepts request from the user and looks for the node to send back the response. Basically it allows the end user to create instances from OpenStack.
The Nova Compute component [26] creates and removes instances through the API?s of the Hypervisor. The compute node handles the lifecycle management through Message Queue and stores the metadata. Nova Scheduler based on the resources allocated to the instance decides on which computer server it should run. Network Service Node handles

the network configurations of the instances, and manages the IP addresses. The Quantum Project has now been replaced by another Project named Neutron. The Neutron Project consists of three main components: Neutron Server, it received the request from API's and directs them to the plugin of that action. Neutron plugins handles the work of network, such as creating subnets, IP addressing and so on. Queue handles the messages between server and agents and MySQL to save the state for a queue. Due to API's extension there is a sophistication of security and quality of service. Horizon is a web based user interface that connects different services together to launch instances through the API?s. The API?s send and receive the requests from the user or the Administrator to the OpenStack services and displays those request to the Horizon. It is a reference point without storing any data except session information.

Additionally the components that offer storage are:

1. Object Storage known as Swift
2. Block Storage known as Cinder

Data storage is the most sensitive part where many engineers are concerned, the obstacle is the downtime when increasing the availability of data. Swift appeared to solve this issue by changing the traditional ways. The Benefits of it are now obvious such as, in case of failure it gives auto recovery, it is highly scalable to store even petabytes of storage, has a better performance by communicating with other nodes to spread the computing job and so on. Block Storage or Cinder stores the virtual machines and the data of those machines. The volumes are attached to their VM instances by giving users interaction to the block storage.

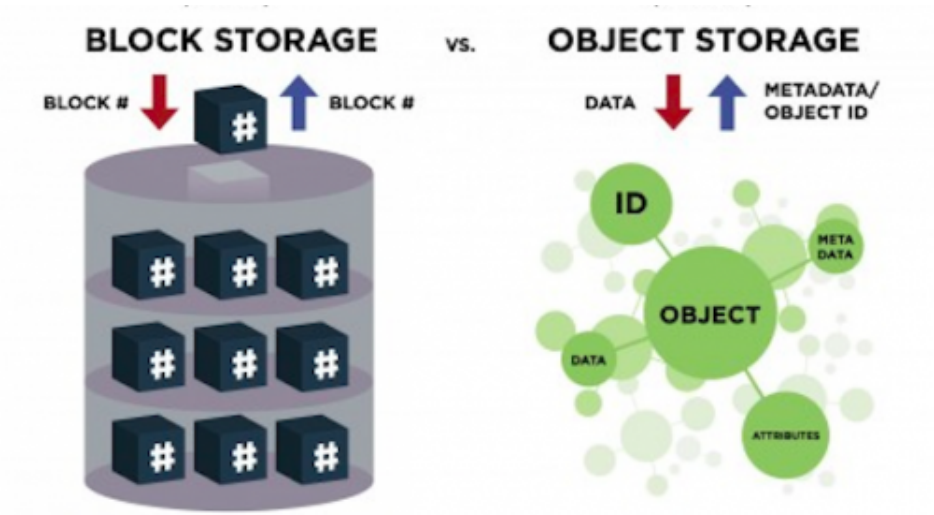


Figure 2.5: Block Storage vs Object Storage [21]

The volumes can be easily detached and reattached again to a different instance because volumes are perpetual. Block Storage drive is used to support back end, these drivers makes it easy to establish access to the block devices of hardware. There are many characteristics of Cinder, it offers volume management and snapshot management, one can easily clone the volumes, it easy to create volumes from snapshots. When deciding for storage management options we ought to know the differences between Object Storage and Block Storage. In terms of Performance Block Storage is the best as well for database storage and volumes. However, in accessing the data in Object Storage it can be done anywhere, whereas in Block Storage only within the Instances. Block Storage is still not perfect because of quality of service, replication and tiering.

Among the share services we have these components:

1. Identity Service known as Keystone
2. Image Service known as Glance

Identity Service basically plays the role of Active Directory of the Windows Servers. Its job is to authenticate and authorize cloud users centrally. The APIs interfere in the middle to ensure that the permitted users have the right to use the requested services. Authentication works in a way that Keystone sends authorization tokens which are given to different services when validated. This same token is further used for authentication and verification.
Glance just like Swift Project are storage systems but they handle different types of data. Glance is used to store and retrieve metadata of the images, this metadata includes information for disk images, format, kernel and so on. We can take snapshots to copy server images and retrieve them again just as templates so that when we want to setup new servers we can do it more quickly. Snapshots for virtual machines is also possible and we can store them in OpenStack Object Storage (Swift).
The Telemetry Project of OpenStack also known as Ceilometer offers the possibility to collect metering data in terms of CPU and network costs. Ceilometer gathers all the measurements needed to establish customer billing and serves as a point of contact for billing systems. It helps the deployers to configure the data type in order to meet the operating requirements. Use cases are other Telemetry requirements whose are diverse and reach beyond a single project.

2.5.4 Community App Catalog

The Community Application Catalog makes it possible to deploy applications that are running independently from the Operating Systems, which means that we can run multiple applications on top of cloud operating system. To make this possible on top of Horizon or the OpenStack Dashboard we need to install certain plugins that offer this service. The concurrent plugin that offer this service is called Murano where one can find the complete applications as well as clustered configurations. The Murano Project helps the Administrators to public cloud applications in the online catalog.
To deploy an application first we need to know the name of the application and then under Application Catalog type the name of the app we want to install, once it appears just click on Download.

2.5.5 Limitations of OpenStack

OpenStack is not meant for a single node installation because there are many disconnected components such as API?s, mysql and so on. Basically it is meant to scale, so without virtualization it will be difficult. However it has been proven from Kasser and Nasim [23] that OpenStack performs much better in a dedicated hardware instead of virtualization. OpenStack for home cloud is another obstacle due to the fact that not everyone who wants it knows how to do the installation. OpenStack does not have an interface which does the automatic installation, instead one must go through all the manual installation and that would be a bit difficult for non IT users. HP has also admitted that the networking component Neutron has many issues. Neutron node is crucial for deploying large Infrastructure as a service, so it had to be combined with a third party plugin to work properly respectively NSX plugin. This was also confirmed by the director of Rad Hat who said ”This is the only Project where we required a third party component to be whole” [27].

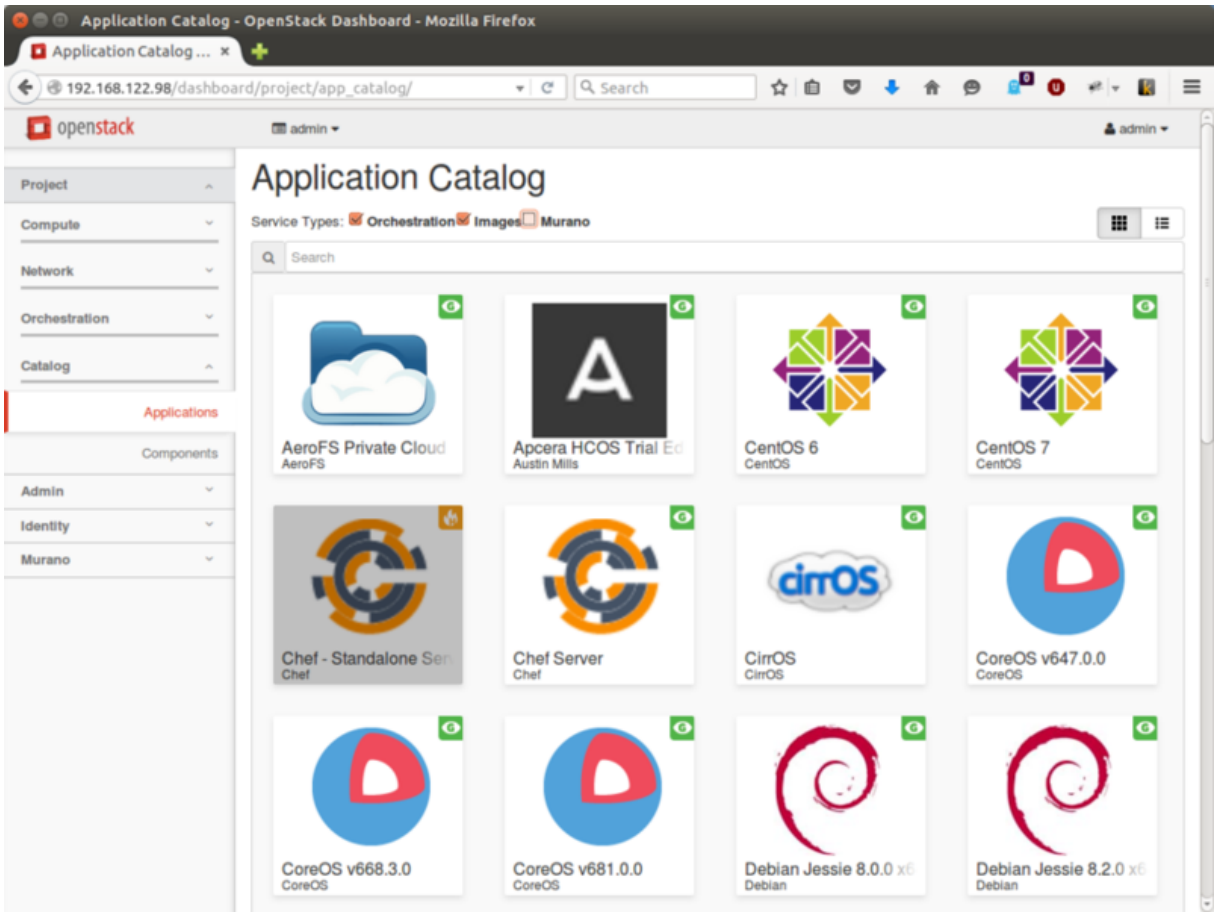


Figure 2.6: The main Interface of Application Catalog

2.6 Reachability of Home Cloud

In the following sections we will present three of the many possible solutions for accessing a home cloud outside of the home network.
The first option is to rent a static IP address for the users to access to their home cloud via Secure Shell (SSH) or other clients shown in Figure 1.7. In the second one shown in Figure 1.8, the user gets a domain name additionally to the static IP address in order to use the domain name for accessing the home cloud. The last solution displayed in Figure 1.9, is opting for a dynamic IP address with a subdomain name or a domain name of a dynamic DNS (DDNS) provider.

2.6.1 Static IP Address

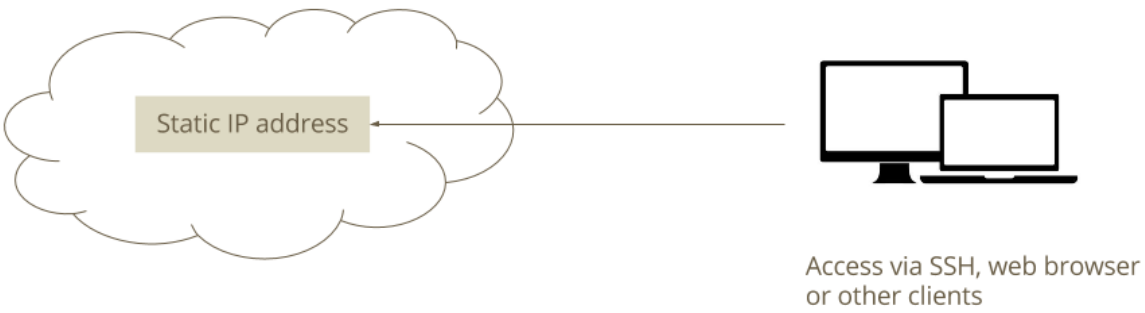


Figure 2.7: Static IP Address

In order to get access to a home cloud via SSH, a web browser or a client like (e.g. ownCloud) outside of the home network, a static IP address is needed. A static IP address has to be rented. There are several providers such as Swisscom, Sunrise, Init7 and iWay in Switzerland that offers static IP addresses, but it is not possible to rent a static IP address only at any of the providers. Which means user needs to sign up for a specific monthly or yearly plan at the provider to obtain a static IP address. For example in case of Swisscom, the user needs to choose the subscription type "Business Internet light" to have an option for a static IP address for CHF 10.- in the first place [7]. The cheapest subscription of this type costs CHF 29.- with a very low bandwith. Thus, the minimal expense for having a static IP address at Swisscom would amount to CHF 468.- per year. At iWay, the offer is similiar to Swisscom. A static IP address costs CHF 10.05 per month which could be rented optionally to one of the subscriptions the provider is offering [28]. The cheapest subscription cost the same, but with a bandwidth five times higher compared to Swisscom. This option costs CHF 468.60, but with both, higher download and upload rate. Another provider with a comparable offer to iWay is GGA Maur. At GGA Maur the static IP address is already included in the Business Internet subscription series "Connect Pro". With the cheapest option at a price of CHF 95.- per month, it makes GGA Maur the most expensive one out of the three providers [29].

2.6.2 Static IP Address and Domain Name

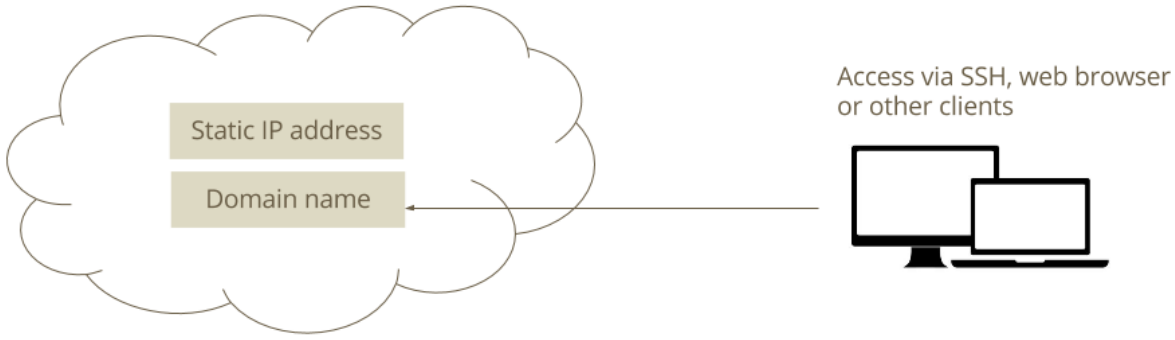


Figure 2.8: Static IP Address + DNS

If a static IP address seems to be inconvenient or confusing for users, a domain name could be rented additionally. So instead of having a static IP address, users can use the domain name to access their home cloud. Although a domain name is not necessary in the setup of a home cloud, it is thoroughly a more elegant solution and might be easier for users to remember than a sequence of numbers. With an annual fee of CHF 15.50 an available domain name of your choice can be rented at switchplus [30]. To make it reachable an activation of the domain name is needed which could be done via an admistration interface [31].

2.6.3 Dynamic IP Address and Domain Name

The third solution is opting for a dynamic IP address in place of a static IP address. In this case a subdomain name of an existing domain or a domain name at a DDNS provider has to be rented. It depends on the provider, if you are able to rent a domain name or just a subdomain name. Similar to the second solution in, a unique hostname is used to access the home cloud instead of the rather confusing IP address. The biggest advantage of this solution is its low cost compared to the other solutions above. For only USD 40.- per year, a remote access (DynDNS) can be rented at Dyn [34]. Because the home cloud

Figure 2.9: Dynamic IP Address + DDNS

does not have a static IP address, an update client of the provider is required to monitor its change of IP addresses. With TTLs such an update can be set as low as 20 seconds. The update client will notify this change to the hostname as soon as the propagation has happened. In short, via this client the hostname will update the current IP address of the home cloud whenever its IP address changes.

2.7 Cost Estimation for Home Cloud Setup

The main cost would be the large up-front cost for the hardware. We can avoid this cost by putting in use a working computer at home that we are not using it. Besides the cost of hardware, the maintenance cost and upgrade cost should be considered carefully later. To have better performance there is always a need for upgrade such as new routers because the value of them will be degrade by the time. To estimate the cost of a running server for OpenStack people may prejudice just by seeing the label of 220 watts on the Power Supply of the Computer. Usually this number tells us the maximum amount in Watts the Power Supply can output. To know the exactly the usage of Power Supply for a computer is very difficult to measure. Different computers have different hardware configurations, for example a high end video card can increase the amount of electricity heavily. Rackspace has developed a web based calculator [19] for estimating the cost of Cloud Environment by adding the servers, databases and add-ons, simultaneously one can specify the bandwidth as well. The drawback of it is that we can only select the US regions for the prices. There is another interesting Power Supply Calculator developed by Micro-Star International (MSI) where can specify exactly the CPU, Graphic Card, Storage, Memory and PCIs. Depending on the efficiency of the Desktop Computer, on average an hour of usage consumes at around 0.2 Kwh, that amound times 24 will give us 4.8 Kwh per day. So from that amount the monthly accumulation can reach at around 145 kwh. In Switzerland for the 2016 it is estimated to be 0.25 Swiss Francs for 1 Kwh [20]. Hence the monthly expenditure for electricity can be at around 36 Swiss Francs. To get the yearly consumption we can use this calculation, Consumption [kWh/year] x electricity price [swiss rapps/kWh] / 100 = Electricity costs [CHF/year]. Regarding the reachability of the home cloud, the cost of shown solution may vary considerably depending on the choice in providers. It is also strongly associated with bandwidth subscription, as static IP addresses cannot be rented separately. Taken from section 2.6, the cheapest option costs USD 40.- per year, whereas other rental options will cost at least CHF 120.- additionally to a chosen subscription type, for example at a provider like Swisscom [7].
With the assumption that the setup of the home cloud is done on a device in stock or that the purchase of any additional hardware is unnecessary, the estimated energy and rental cost for setting up a home cloud would amount to approximately CHF 500.- per year.

2.8 Security

Cloud computing enables the user to store the data online at the same time it faces security threats and challenges of protecting the user data from unauthorised users. Many security controls has been suggested by the standards and the governing bodies but due to the challenges of low availability, less user convenience and robust infrastructure it has not been fully implemented. So therefore to implement the data protection in home cloud based on categorisation of data determined by users.
Data in the cloud is subjected to various types of attacks such as man in middle attack, brute force attacks, sniffing which raises security concern for the users. Cloud service provider has implemented some security measures in home cloud are two factor authentication, encryption etc making access to the user data more difficult for attackers. Encryption cannot fully guarantee to protect user data and also to encrypt the whole data they need high robust infrastructure which is highly expensive hence its not the best option.
The basic data protection mechanisms employed in cloud storage are.

- Authentication: The user has to prove its identity to the server or client. To identify the user it consist of username or user ID and password to confirm the credential is valid
- Authorization: Here it determines if the user has permission to use a resource or access a file. It provides an authenticated user the ability to access the application, data set, data file. It assigns rights to the users to use, modify or delete items.
- Encryption: Is the process of transforming data so it makes it unreadable to anyone without having decryption key. This mechanism is employed to provide security measures to protect data at rest, in process and in transit.

Security levels to protect data in cloud storage can be implemented such as protected, sensitive and top secret based on the data determined by the users as shown in the Figure 2.10

- Protected level: We usually have data that is for public and free distribution, which are not critical to the user so we just going to employ single factor authentication such as password protection. So therefore for the encryption as for the Encryption you should use a SSL (Secure Socket Layer) and TLS (Transport Layer Security) for the data transfer to create a secure tunnel. Ones it reaches the cloud storage it is protected using 256-bit AES encryption at rest.
- Sensitive level: We have data that would not have a severe impact on the user if they are lost but not for the public view. Here we could use a multi factor authentication (a two-step verification). As for the encryption it is sufficient to use the same one as on the protected level.
- Top-secret level: We have data that is classified as confidential It is recommended to also use a multi-factor authentication. Here the encryption is more complex. Above from using SSL/TLS, 256-bit encryption the data in use is protected using 256-bit AES and RSA encryption [13].

  In general the different security levels, the authentication and encryption employed in each level can be seen in Figure 2.11

Considering the major security issues of cloud computing, some safety measures for cloud computing are

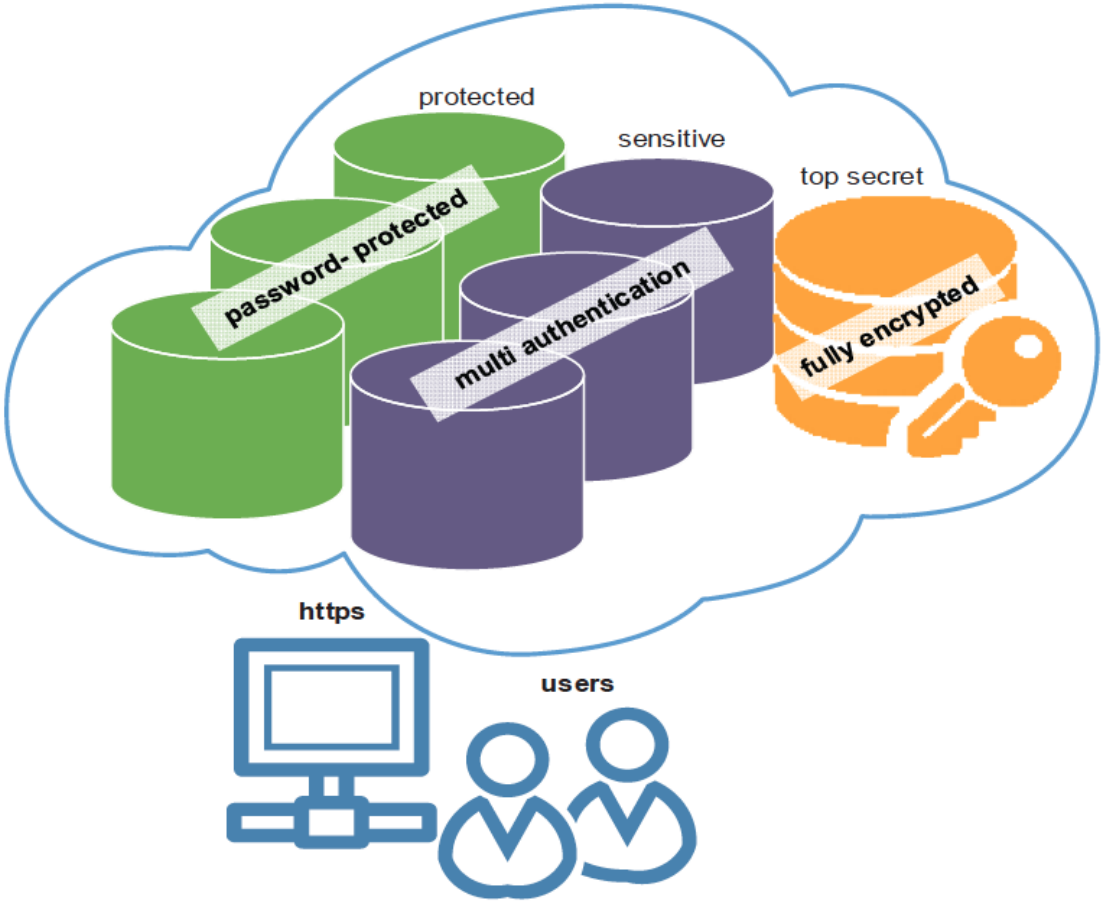


Figure 2.10: Security classification to Protect Data in Cloud Storage [35].

- Reinforce anti attack capability: Deploying anti virus software and firewalls in the cloud
- File encryption: By encrypting the file it protects the user data from unauthorised users.
- Protecting API key: Protecting the API key to prevent illegal users to access the cloud data
- Data Backup: Users should back up some of the most important and sensitive data to protect it from stolen or tampered by hackers so that it can restored by the backup data [14].

2.9 Security Risk

Security risk associated with cloud computing depends on number of factors such as information assets, cloud architecture, and security control.
List of security risk in cloud computing are:

- Privileged user access: Cloud user data processed outside the enterprise brings with it inherent level of security risk. To control the access of data from the other cloud users and from the third party supplier, the cloud service provider should employ the mechanism of encrypting the data before they put in the cloud. The user might not give access to their sensitive data to the provider although they might treat the data confidentially, this problem is eliminated in private cloud, since the user itself

Security Level	Authentication	Encryption
Protected	Single Factor	SSL, 256-bits AES
Sensitive	Multi-Factor	SSL, 256-bits AES
Top Secret	Multi-Factor	SSL, 256-bits AES, RSA, Filename Encryption

Figure 2.11: Security Levels [35].

will act as the cloud provider. And also most of the time, the user is not happy with the encryption technology the cloud provider is using, so therefore by launching the private cloud, the user can employ the encryption technology of their choice to protect their data.

- Data location and segregation: In case of public cloud, the cloud users may be under confidence, regulatory or contractual obligations to ensure that data is held, processed and managed in a certain way by the cloud provider. The number of security risk in this situation might be the cloud provider might disclose the data to a third party, environmental hazards such as earthquakes, flooding affecting the security of cloud user data, hyper inflation affecting the providers service and personal conditions.

- Data disposal: In public cloud data deletion and disposal is a risk, as the cloud provider might keep multiple copies of the data to guarantee high availability of data to the user, but this problem is eliminated in case of private cloud.

- e-investigations and Protective monitoring: Implementing protective monitoring in the cloud possess challenges to both providers and cloud customers to access the location of physical data when high number of providers are involved. There is risk of insider threads and attacks on the cloud and requires e-investigations and monitoring. For public cloud users there is limited options to invoke electronic investigations procedures within the cloud. As the infrastructure is not owned by the user they cannot effectively deploy monitoring systems, so public cloud must always rely on the cloud service provider to support investigation. In private cloud as user own the infrastructure it is easier to implement protective monitoring on their cloud [2].

- Assuring cloud security: Currently there is no common cloud computing security standards for the public cloud users to assess their providers. They cannot easily assure the security of system that they do not directly control without using the Security-level agreement and having the right to edit security controls within their agreements. In private cloud the users don't have this kind of risk at all, as the security is in the hand of the users setting up the private cloud.

2.10 Advantages and Disadvantages of Home Cloud

As there are some advantages of setting up the home cloud, there are also some disadvantages related to it.
Some of the advantages of setting up the home cloud are:

- Greater control: Home users have more control over their data. The administrator is in charge of monitoring and maintaining the data giving users complete oversight of their data. But in case of public cloud third party providers are in charge of the data systems, many users feel as if they don't have enough control over their personal data with a public cloud service.

- More security: Storing confidential and sensitive information in the home cloud network is more secure than storing it locally [16]. In home cloud data is encrypted both during transmission and also at rest making unauthorised users difficult to access the file and will also assure the user of data being in their control. It is just that home cloud users will feel the data is more secure by having it reside in-house.

- Customizable: Network performance and storage performance can be specified and customized in the home cloud since it is owned by the home users.

Some of the drawbacks are:

- Higher Cost: In order to read and store large files in cloud computing the network connection of high bandwidth is required, which may be available for individual home cloud users but cost will be the constraint to purchase high bandwidth connection. So home users suffers from a trade-off between the network bandwidth and and the cost that the network sets on its resources.

- Quality of Service: In case of corporate company's setting up the private cloud will enable powerful servers to manage the cloud to satisfy the clients and customers, but in case of home cloud user as cost is the constraint the quality of service will be low as compared to the private company's and this will ultimately degrade the quality of experience for the clients and for the customers [15].

- Security Issues: Corporate company deploy private cloud inside the firewall to ensure greater security. But for personal users or for small business cost is the constraint they cannot ensure high quality security as like a corporate company's. In that case home network will be subjected to network attacks such as distributed denial of service, man in middle attack.

2.11 Evaluation and Conclusion

Nowadays, there seems to be a big misconception about clouds, as not every so-called cloud is actually based on cloud computing. Nevertheless, cloud computing has emerged from previous concepts and has gained importance. Clouds has become more popular in recent years, but in the midst of its rise in popularity, the discussion about its security has also become more important. Public clouds often lack in security and privacy where private clouds is the answer to this problem. Unfortunately, these private clouds are often very costly, therefore only bigger companies or institutions are able to afford them. For private persons, home cloud as a special case of private clouds, is feasible a solution, which offers both the benefits from cloud computing technology and security.
With Open Source cloud software like OpenStack, a home cloud can achieve an architecture similar to the clouds on the market. As OpenStack is very extensive with its

numerous features and components, it is often used for setting up private clouds in enterprises. However, it could also be used to set up a home cloud.
There are many arguments for using a home cloud instead of a public one, but also many against it. The main argument for using a home cloud is security. Storing data in a home cloud is more secure, especially if the data is highly sensitive. Also the users are not obliged to security regulations from the provider, but they can technically choose their own level of security. By owning their cloud, users do not have to worry about a third party accessing their data or where their data is located. Users are able to choose their own infrastructure and therefore also what kind of protective monitoring they want to implement on their home cloud. But there are also some drawbacks. Choosing OpenStack as the cloud software for a home cloud could be challenging without profound knowledge due to its large number of features and components. The simpler option to OpenStack would be onwCloud which provides similar functionality to the users. The maintenance of the home cloud is as important as the setup, particularly if there are more than one person using the same home cloud. At least one person has to be responsible for maintaining the home cloud as well as for trouble-shooting if something goes wrong, which could be time-consuming. Because cost is a constraint, the quality of service might not be as satisfying as in pubic or private clouds.
The reason why people do not setup or use a home cloud more often is inconvenience. It is given that public clouds provide incredible convenience to their users. Most user are unaware of the security breaches in public clouds or do not care enough about it to look for an alternative solution. Even if they care about the security and have the necessary knowledge, they are not absolutely willing to take the time to setup their own cloud. Some may even say that it is too inconvenient for them to do so. The setup of a home cloud is also associated with additional expenses like energy cost, rental cost and it requires time from the person who does the setup. Users tend to choose the more convenient existing public cloud solutions rather than to build a more secure private one. Alternatively they could use a NAS device or a server for storage and file sharing instead, which almost serves the same purpose as a home cloud.
Regardless of its low popularity, home cloud is a good supplement to existing solutions of cloud computing.

Bibliography

[1] ITU-T: Privacy in Cloud Computing; ITU-T Technology Watch Report, March, 2012, http://www.itu.int/dms_pub/itu-t/oth/23/01/T23010000160001PDFE.pdf. Last accessed December 2015.

[2] J. Sen: Security and Privacy Issues in Cloud Computing, Innovation Labs, Tata Consultancy Services Ltd., http://arxiv.org/ftp/arxiv/papers/1303/1303.4814.pdf. Last accessed December 2015.

[3] S. Goyal: Public vs Private vs Hybrid vs Community - Cloud: A Critical Review, I.J. Computer Network and Information Security, March, 2014, 20-28.

[4] S. Viskas, K. Grundat, M. Vishnu, K. Prashant: Private Vs Public Cloud, International Journal of Computer Science & Communication Networks, Vol. 3(2), 79-83.

[5] P. Mell, T. Grance: The NIST Definition of Cloud Computing, NIST Special Publication 800-145, September, 2011, http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf.

[6] Z. Pantic, M. A. Babar: Guidelines for Building a Private Cloud Infrastructure, Tech Report TR-2012-153, University of Copenhagen, 2012.

[7] Swisscom: Business Internet light, https://www.swisscom.ch/de/business/kmu/internet/internet-im-buero/business-internet-light.html. Last accessed December 2015.

[8] D. Hodanic, N. Vrkic, M. Tomic: Data Storage and Synchronization in Private Cloud, Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2015 38th International Convention on, 476 - 480, 10.1109/MIPRO.2015.7160318, May, 2015.

[9] V. Henry: Benefits of Private Cloud over Public Cloud, June, 2015, http://www.datamation.com/cloud-computing/benefits-of-private-cloud-over-public-cloud-2.html. Last accessed December 2015.

[10] Forrest Stroud: Private Cloud Project, March, 2011, http://www.webopedia.com/TERM/P/private_cloud_project.html. Accessed December, 2015.

[11] Y. Chou: What is Private cloud, March, 2011, http://blogs.technet.com/b/yungchou/archive/2011/03/21/what-is-private-cloud.aspx. Last accessed December 2015.

[12] A. Gajbhiye, K. M. P. Shrivastva: Cloud computing: Need, enabling technology, architecture, advantages and challenges, Confluence The Next Generation Information Technology Summit (Confluence), 2014 5th International Conference, 10.1109/CONFLUENCE.2014.6949224, 2014, Sept, 1-7.

[13] F. Yahya, R. J. Walters, G. B. Wills: Protecting data in personal cloud storage with security classifications, Science and Information Conference (SAI), 2015, 10.1109/SAI.2015.7237241, 2015, July, 838-843.

[14] J.-J. Wang, S. Mu: Security issues and countermeasures in cloud computing, Grey Systems and Intelligent Services (GSIS), 2011 IEEE International Conference on, 10.1109/GSIS.2011.6043978, 2011, Sept, 843-846.

[15] S. Bhattacharyya, M. Sen, S. Agarwal: Battling the Fear of Heights: Pros and Cons of the 'Personal Cloud, Communication Systems and Network Technologies (CSNT), 2012 International Conference on, 0.1109/CSNT.2012.159, May, 726 - 730, 2012.

[16] K. Nielsen: Top Ten Advantages of Using Online Storage Services, March, 2012, `http://online-storage-service-review.toptenreviews.com/top-ten-advantages-of-using-online-storage-services.html`. Last accessed December 2015.

[17] R. McCarty: Virtualization, September 2014, `http://searchservervirtualization.techtarget.com/tip/Paravirtualization-explained`. Last accessed October 2015.

[18] OpenStack, `http://docs.openstack.org/juno/install-guide/install/apt/content/ch_overview.html`. Last accessed November 2015.

[19] Rackspace, 2015, `http://www.rackspace.com/calculator`. Last accessed December 2015.

[20] Swissgrid Tarrifs, 2015, `https://www.swissgrid.ch/swissgrid/en/home/experts/topics/grid_usage/tariffs.html`. Last accessed December 2015.

[21] Y. Porter de Len, T. Piscopo: Storage, 2014 `http://www.druva.com/blog/object-storage-versus-block-storage-understanding-technology-differences`. Last accessed December 2015.

[22] OpenStack 2014 `https://wiki.openstack.org/wiki/Main_Page`. Last accessed October 2015.

[23] R. Nasim, A. Kassler: Deploying OpenStack: Virtual Infrastructure or Dedicate Hardware , Publisher "Research Gate", 2014, 3-7.

[24] R. McCarty: Virtualization 2014, `http://searchservervirtualization.techtarget.com/tip/Paravirtualization-explained`. Last accessed November 2015.

[25] Wikipedia 2014, `https://en.wikipedia.org/wiki/Kernel-based_Virtual_Machine`. Last accessed November 2015.

[26] O. Khedher: Mastering OpenStack, Packt Publishing, ISBN 1784395641, July 2014.

[27] J. Clark, Cloud brains at HP, Red Hat, Piston, spill beans on the weak link in the OpenStack cloud, May 2014 `http//www.theregister.co.uk/2014/05/13/openstack_neutron_explainer/`. Last accessed October 2015.

[28] iWay: Internet, `https://www.iway.ch/privatkunden/stadtnetz-ewz`. Last accessed December 2015.

[29] GGA Maur: Business Internet, https://www.gga.ch/zuerich/produkte/internet/business-internet/. Last accessed December 2015.

[30] switchplus: Domain Services, https://switchplus.ch/en/domain-names/. Last accessed December 2015.

[31] Switch - Internet Domains, https://www.nic.ch/reg/cm/wcm-page/faqs/manage.jsp?lid=en#activate. Last accessed December 2015.

[32] simple talk: Private Cloud, What Is It and Why Do You Need It?, https://www.simple-talk.com/cloud/development/private-cloud,-what-is-it-and-why-do-you-need-it/. Last accessed December 2015.

[33] simple talk, http://blogs.technet.com/b/yungchou/archive/2011/03/21/what-is-private-cloud.aspx. Last accessed November 2015.

[34] DynDNS: Remote Access (DynDNS) Free Trial, http://dyn.com/dns/dyndns-pro-free-trial/. Last accessed December 2015.

[35] F. Yahya, G.B.Wills, R.J. Walters: Protecting data in personal cloud storage with security classifications, Science and Information Conference (SAI), 2015, 10.1109/SAI.2015.7237241, July 2015.

Chapter 3

VANET Services on the Road: A Contrary Perspective of VANET Services

Arlind Bejta, Jens Hinkelmann, Sascha Gabler

This paper provides a general overview of the VANET networking, samples of safety and non-safety applications and discusses a contrary perspective related both to safety and to non-safety applications. Moreover, this paper addresses the reasons that VANET technology has not been implemented for non-safety applications.

Contents

3.1 Introduction **55**
3.1.1 VANETs Analysis and Definition 55
3.1.2 Technical Specifications and Communication Standards 57
3.1.3 VANETs Characteristics 57
3.2 Assessment of a real life VANET-application **58**
3.2.1 Issues 58
3.2.2 Approach 60
3.3 Socio-economic aspects of VANET **64**
3.3.1 Successful market introduction 64
3.3.2 Controversy of new safety technologies 66
3.4 Potential customer reception **67**
3.4.1 Safety applications 67
3.4.2 Non Safety applications 68

3.1 Introduction

The growing number of world population carries a higher demand for vehicular transportation. As the number of vehicles increases, the risk of accidents, traffic congestion, and last but not least, our environment is at risk. According to the Global Status Report on Road Safety, there are around 1.24 million of road-related accident deaths and 20 to 50 million of people having non-fatal injuries as a result of road accidents. These numbers correspond to the eighth worldwide leading cause of deaths and are the leading cause of death for the group age between 15-29 years old [5]. The report additionally includes global information about the current restraints and laws that countries have implemented on their roads to increase safety. However, "only 59 countries, covering just 39% of the world's population (2.67 billion people), have implemented an urban speed limit of 50 km/h or less and allow local authorities to reduce these limits."[5].
Considering this fact, wiser solutions have to be made so that the lives of the people will not be exposed to such risks. The European Commission has issued a "Road Safety Programme" that targets to lower the road related deaths. The programme was released on 2011 and it will continue up to 2020. According to the report, approximately 30'000 people have died on the roads of the European union, for comparison sake, that is the size of a medium town in Europe. In addition, the proportional ratio is that for each death occurred on European Union roads, there are 4 accidents resulting in invalidity [6].
Moreover, in Switzerland there are over 5.8 million vehicles, and 75% of the vehicles are privately owned. In terms of statistics, every second citizen owns a car [7]. As the consequence, there is a great chance that road travelers can be involved in an accident. According to the Swiss Council for Accident Prevention, there were more than 18'000 people who were severely or slightly injured and 180 fatalities as a result of road accidents [8]. Thus, the continuous and increasing problem with the traffic jams, air pollution, and accidents have led the researchers and large automotive corporates to find solutions to increase the road safety and management. Such a solution includes a Vehicular Ad-hoc Network (VANET) where the objective is to have higher safety on roads, increase the road efficiency and provide additional applications and services for the drivers and passengers. Therefore, the collaboration among researchers, automotive industry, shareholders and the government provides an opportunity for both commercial and scientific benefits.

3.1.1 VANETs Analysis and Definition

A typical Wi-Fi network consists of a single or many devices connected to an access point, where an access point is then connected to a wired connection. The slowest connection happens between the device and the access point. Similarly, to establish communication between cars, a network has to be created so that the information could be processed from one vehicle to another vehicle. However, a vehicle might be located far from an access point. During this situation, an intelligent solution is to provide exchange of information among cars by turning the cars into routers [12].
The Vehicular Ad-Hoc Network (VANET) is an ad-hoc network. An ad-hoc represents a connection that is available only when needed, thus eliminating the need for a permanent connectivity. Moreover, the ad-hoc connection will be enabled when a vehicle is not within the range of a Road Side Unit (RSU) coverage. Moreover, in Figure 3.1 are shown possible combinations of communication methods. Such methods include vehicle-to-vehicle communication, vehicle-to-roadside communication, and inter-roadside communication. During the communication, the nodes transmit their position to other nodes or to RSUs by utilizing GPS.

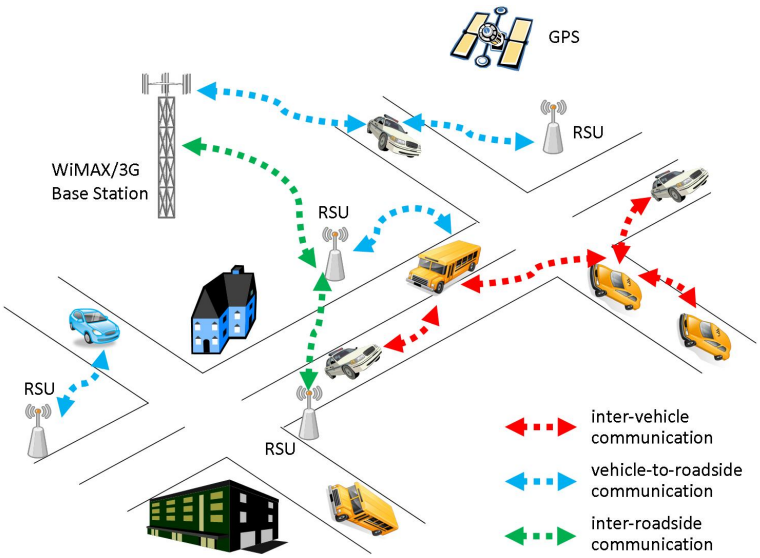


Figure 3.1: Vehicular Ad-Hoc Networks (V2V, V2RSU, I-RSU) [9]

3.1.1.1 Vehicle-to-Vehicle or Inter-Vehicle Communication

The communication among vehicles is an important factor in this network. It establishes communication between two or more vehicles (nodes), such as cars, buses, emergency vehicles, trucks, via a short or long range communication, and in this way obtains the necessary information for surrounding environment [10]. The communication is set possible through two types of broadcasting: "naive-broadcasting" and "intelligent-broadcasting" [11]. Naive broadcasting allows communication among vehicles in such a manner that only messages that are send from vehicles in front are received, and messages send from vehicles in back are ignored. However, this type of communication sets a large number of messages and creates congestion resulting in wasted messages. On the other hand, the intelligent-broadcasting assures that only emergency messages are sent. This is allowed by having the vehicle's forward only messages that have not been received from the car behind [11].

3.1.1.2 Vehicle-to-roadside communication

VANETs have access to the Internet through a wired infrastructure which is generally located in the side of the road. Hence, it is called a Road Side Unit (RSU) and it can be a traffic light, a traffic sign, and up to an antennae [11]. Communication is set via the RSU by hopping messages to the first vehicle in its range. Then once other vehicles are in range of the first vehicle, the previous messages will hop to those other vehicles. By repeating this process, the RSU can send relative messages to other vehicles or even to a traffic analyzer. Messages that are sent can be information provided by the RSU to optimize the ongoing traffic. Therefore, using this infrastructure brings benefits on controlling the traffic to bring positive implications such as optimization of fuel consumption and lower emissions [11].

3.1.1.3 Inter-Roadside Communication

The inter-connection among roadside units and other nodes allows wider communication. Such information sent from RSU-RSU could provide broader view of the road conditions. In addition, a RSU connects to a base antennae and interchangeably transmit information. Such information could at later time be hoped from an RSU to a moving node to provide near by services [11].

3.1.2 Technical Specifications and Communication Standards

3.1.2.1 Dedicated Short Range Communication (DSRC)

Dedicated Short Range Communication is a protocol that ensures standardization of frequencies used by VANETs [12]. Channels are divided into:

- one way and two way communication channels and,
- short-to-medium range wireless communication channels specifically designed for automotive use [12].

In addition the frequency of communication is set at 5.9 GHz and bandwidth allocation of 75 MHz. The channels further on are divided into control channels and service channels in order to prevent congestion. In addition, there are seven channels with 10 MHz and are introduced in the following format: 172, 174, 176, 178, 180, 182, 184. The format set for the United States standards occupies channel 178 as a control channel while the rest of the channels are allocated as service channels [13].

3.1.2.2 IEEE 802.11p Wireless Access for Vehicular Environment (WAVE)

Wireless Access for Vehicular Environment is based on the IEEE 802.11a architecture and it provides a set of protocols and standards for security when transmitting messages. Road Side Units and On-Board Units are the main devices that use these services. The Figure 3.2 shows the distinct view of WAVE and DSRC [13].

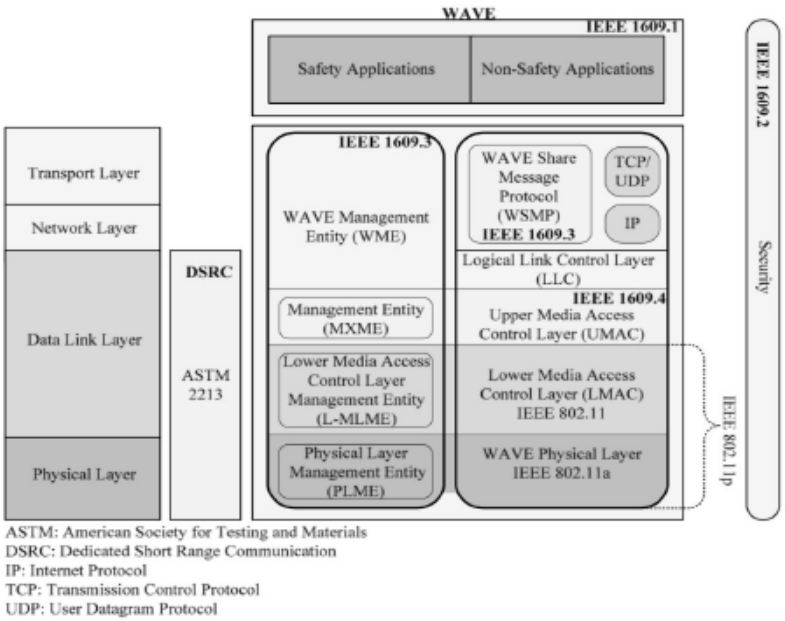


Figure 3.2: Communication standards DSRC and WAVE under OSI model [11]

In addition, Figure 3.2 also provides the WAVE Share Message protocol (WSMP) that falls in between Transport layer and Network layer and under Application layer. This protocol ensures that every Basic Safety Message (BSM) must be authenticated before being received. A BSM is sent every 300 ms and at any given time hundred of nodes receive messages [13].

3.1.3 VANETs Characteristics

There is a significant resemblance between Vehicular ad-hoc networks (VANETs) and Mobile ad-hoc networks (MANETs). However, there are distinguished features that characterises VANETs from MANETs. These features inlcude [21]:

- *Large dynamic topology* The communication is dependant on the speed of the vehicle and the transmission distance. A large dynamic topology is hard to predict.
- *Periodic disconnections* The high speed of the moving nodes creates large dynamic topology. In the same time, it also creates the necessity of the roadside units (RSUs). RSUs need to be distributed very frequently to avoid disconnections. On cases when a vehicle is not in range of other vehicles, messages will be transmitted to RSUs, thus a disconnection could occur if RSUs are not distributed frequently [21].
- *Predictability* Predicting a vehicle position and their corresponding routes is hard. Predictability is based on already existing road maps models.
- *Transmitting constraints* Transmitting a message on emergency situations is highly important. In addition, there is a limit while transmitting a message, up to 1 Km [13].
- *Energy and storage* Engergy supply and storage related problems are not an issue in VANETs.

3.2 Assessment of a real life VANET-application

To facilitate the handling of VANET as a subject, we discuss a specific VANET-based product concept, based on existing ideas. This product aims to increase car drivers' awareness of their immediate environment. It aims to warn them about approaches by emergency vehicles and other relevant obstacles on their route.

3.2.1 Issues

Various laws in the European Union, Switzerland and also other countries focus on limiting the distraction of vehicle drivers, as driving a car requires a driver's full attention. For example, in Switzerland a driver has to make sure by law, that his/her full attention goes towards driving the vehicle. Still, from 1992 to 2010, inattention and distraction have been contributing factors in 17% to 27% of traffic accident casualties in Switzerland [2]. The car industry already puts great effort into the development of systems that facilitate the driver's task to understand his/her environment. They are referred to as Driver Workload Management Systems. Adaptive cruise control is probably the most famous Driver Workload Management Systems. It helps drivers to adapt their speed to have reasonable security distances to vehicles in front of them. After turning these systems on they often adapt the speed without requiring additional input from the driver. Lane assistants, as another example, are systems that warn drivers when visual sensors detect that the car is crossing a road marking onto another lane even though the driver did not previously set the respective indicator. Up until this point in time, Driver Workload Management Systems implemented in production vehicles focus on the immediate environment as it is perceived by on-board systems and sensors. However there are cases where - often due to high speeds involved - major security threats to a vehicle's passengers can not be recognized fast enough by systems mounted to that respective vehicle to prevent unfavourable outcomes, e. g. accidents; and therefore, automatic and quick communication with a near-by system is presumably necessary to implement a Driver Workload Management Systems to deal with such a threat.

3.2.1.1 Emergency vehicles

In many countries today, under appropriate circumstances, emergency vehicles like ambulance, firetrucks and police cars are empowered to execute high risk maneuvers in a regular road traffic environment. This generally follows the assumptions that these vehicles are easily detected by other traffic participants due to prominent visual and audible signals. In all well known cases, these signals are limited to flashing lights and sirens emitted from the vehicles. Many years of experiences with this kind of emergency vehicle usage show that the effectiveness of these signals is limited. Due to the limited reception range and commonly present blind spots, civil road users are often surprised by the appearance of emergency vehicles crossing their way. This can easily result in congestions, slowing down traffic and therefore also the emergency vehicle, thus hindering the emergency driver's mission. Alternatively, improper reactions by road users can cause accidents that involve either the emergency vehicle directly or only civil road users. Both cases, i. e. the chance for faster mission accomplishment and the possibility to increase traffic security motivate the goal to increase the awareness of civil road users for emergency vehicles rushing to a place of action.

3.2.1.2 Obstacles

Pile up accidents are an extremely dangerous form of traffic accidents. They often include multiple vehicles and each vehicle possible experiences multiple collisions. They typically occur in low visibility conditions and involve high speeds, as under these conditions, drivers often overestimate their capabilities and underestimate their reaction time. Then, when a driver is surprised by some obstacle on the road, e. g. a car that broke down or the rear end of a traffic jam, he/she might have to brake drastically or commit another unforeseeable reaction. This can lead to chain reactions, as following drivers will be surprised by the abruptly moving vehicle in front of them, possibly leading to multiple and heavy collisions (see figure 3.3).

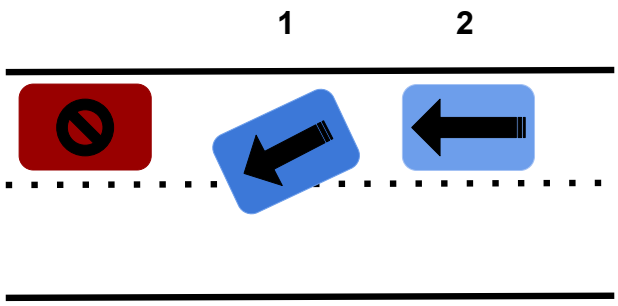


Figure 3.3: Vehicle 1 conceals an obstacle from vehicle 2.

There already exist distance control systems in production vehicles which try to aid drivers with the recognition of obstacles in their way. These systems usually work by means of radar and sonar waves. However, these systems fail when the speed differences between the host car and the obstacle are too big or a slow obstacle is hidden behind a vehicle carrying a speed appropriate for its distance.

One approach to warn vehicle drivers of obstacles that are out of reach of the vehicle's sensors is by transmitting the information over automotive navigation systems. While this solves the problem of recognising obstacles over larger distances, the reaction time is highly limited, as today's automotive navigation systems receive routing information from central systems that have to collect and distribute a lot of information from different locations and also work over slow networks. The limited reaction time is especially of concern, when obstacles developed recently. Possible cases are malfunctioning cars, collisions and situations involving wildlife or pedestrians.
This motivates the goal to aid drivers with additional means to react to obstacles on the road. A new solution should be able to provide drivers with information of obstacles further away than distance control systems based on sensors mounted to the affected vehicle could possibly measure. At the same time, it should unequivocally propagate respective information faster than current automotive navigation systems' traffic and obstacle warnings.

3.2.2 Approach

Warning drivers of emergency vehicles is "often cited as a VANET use case"[1]. VANET has two dominant properties that speak for its use in the application to be developed:
Location dependent: For the issues at hand, the information regarding a driver's environment is only interesting at a certain location; information regarding obstacles and approaching emergency vehicles only have to be propagated to vehicles that are in near to medium distance. They are not of interest for vehicles that are too far away to actually be at risk of coming into contact with them. VANET technology, as it is ad-hoc, is inherently suitable for applications with this requirement.
Latency critical: To sensibly increase drivers' awareness and assist them in their interaction with their environment, a Driver Workload Management System should be as quick or quicker to react to changing circumstances as the driver would be capable of being. As mentioned in regard to automotive navigation systems, centralised services that work over GSM networks to provide traffic warnings are usually overwhelmed with the task of providing location specific data in real-time. VANET technologies are capable of providing proper real-time communication over short to medium distances, which makes them fit perfectly to the requirement of low latency.

3.2.2.1 Requirements

The issues to be solved as well as further considerations amount to the following list of requirements[1]:

- *Versatility:* The system should be designed to handle various different scenarios in individual ways. Consequently, different scenarios should be clearly distinguishable by the respective warnings and alerts.

- *Timing:* As mentioned above, warnings and alerts have to reach the driver quickly enough so he/she can react appropriately to the respective scenario.

- *Relevance:* To ensure a high responsiveness of vehicle drivers to a warning system in general, it is important to only provide considerably relevant information. This prevents the driver from getting used to false positives and conclusively becoming dulled and possibly ignore future alarms in suspicion of it being a false one.

- *Dependability:* Similar to the requirement of relevance, it is important to avoid false negatives, i. e. situations when an alert would be sensible but does not occur. This could, instead of improving the situation, cause additional confusion, as a

driver might not trust his own eyes when the warning system does contradict his perception.

- *Security:* As common with means of wireless communication, in a VANET application, to propagate warnings, the origin of warnings has to be authenticated to prevent abuse of the system. Otherwise, malicious agents could cause chaos byfalsely making half of the drivers on a road suddenly halt for a non-existing obstacle or "illegitimately send emergency vehicle warning messages to gain a driving advantage "([1], III.A).

- *Privacy:* VANET technologies in general comes with the concern of trackable devices and there already exist mechanisms like dynamic MAC addresses[3] to deal with this. Additional privacy concerns come to mind when thinking of the information provided by emergency vehicles. Specifically, "the final destination of an ambulance should not be included in warning messages to protect privacy of the patient"([1], III.A) and similar concerns exist for other types of emergency vehicles.

- *Coordination:* It is important to prepare for situations where multiple emergency vehicles happen to be in the same area. In these situations, contradictory warning messages have to be prevented.

- *Usability:* To finally ensure the effectivity of the system, warnings and alerts have to be provided in a useful and sensible manner. In this regard, the presentation of warnings has to happen in a quickly and easily understandable way while it at the same time "should not induce stress"([1], III.A).

3.2.2.2 Features

When equipped with the system, drivers can rely on the following features:

- *Emergency vehicle warning([1], III.E, III.F):* When an emergency vehicle is approaching, the warning system in a vehicle that receives a respective signal over VANET should: simply inform the driver of the vehicle's presence, when no action by the driver is required but the emergency vehicle is probably perceived by the driver visibly or audibly. actively warn the driver about the vehicles approach and instruct him how to take action.

- *Obstacle warning:* When vehicles experience dangerous conditions that are recognised by their sensors, e. g. brakes lock in an emergency halt or the engine management system recognises an engine failure, other vehicles underway to the respective location receive warning messages similar to the ones that are triggered by emergency vehicles

The warnings will be a combination of visible and audible signals. When possible, warnings should be deployed 30 seconds before action has to be taken([1], V). Actions that drivers can be requested to do will be ([1], III.F):

- Stop at an intersection to let an emergency vehicle pass. See figure 3.4.

- Drive right for an oncoming emergency vehicle. See figure 3.5.

- Drive right for an overtaking vehicle. See figure 3.6.

- Form a corridor for an approaching vehicle. See figure 3.7.

- Slow down for an obstacle. See figure 3.8.

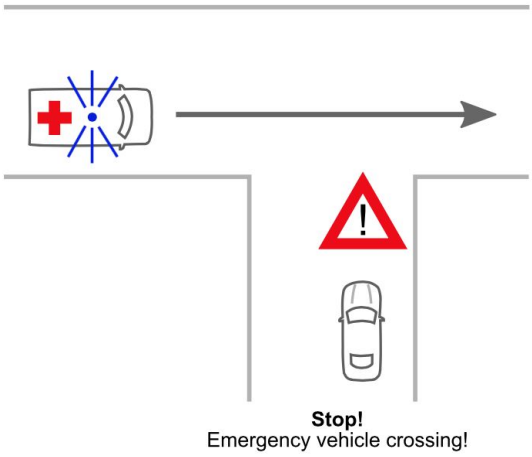


Figure 3.4: Intersection warning[1]

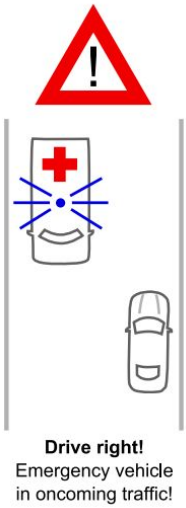


Figure 3.5: Oncoming EV warning[1]

3.2.2.3 Technical details

For approaching emergency vehicle warnings, messages are broadcasted from the emergency vehicle over VANET. For these messages the necessary information for all receivers is required to decide on further actions to take:

- current position of the emergency vehicle
- current velocity vector of the vehicle
- planned route of the emergency vehicle

The required data can be acquired over GPS location systems and on-board diagnostics (OBD). To conform with requirements of privacy, the planned route is never disclosed with the final destination of the emergency vehicle. Instead, the next 30 seconds (warning period) of the route plus a buffer to compensate for propagation delay will be sent. The signal will be propagated by the following set of rules:

- If a receiver of the message is approaching a route that will be used by the emergency vehicle in the next 30 seconds according to the message, an active warning will be shown to the driver and the message is propagated along the route.

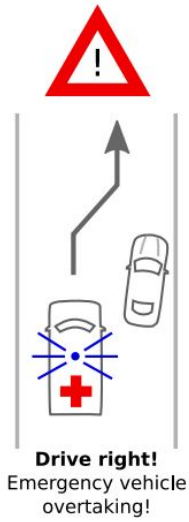


Figure 3.6: Overtaking EV warning[1]

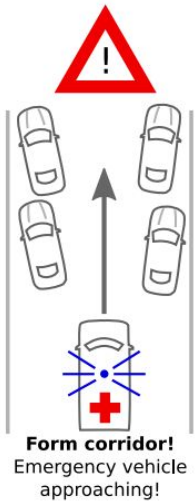


Figure 3.7: Prompt to form a corridor[1]

- If a receiver of the message is in close range to the emergency vehicle but no action has to be taken (e. g. vehicle is on another route but audible/visible), information is provided to the driver and the message is propagated along the route.
- If a receiver is out of visible/audible range and not near the route of an emergency vehicle, the message is only propagated

Messages only have to be propagated for 30 seconds, after which the message is factually useless.
They will be sent over the channel 184 of WAVE protocol, as "the Channel 184 shall be used for the high-power long-distance communications that support public safety applications"[12]. Senders are placed in emergency vehicles and triggered together with sirens. For the warning about obstacles, messages have to be broadcasted from regular traffic participants' vehicles. These messages further contain position and velocity of the vehicle and also information of what triggered the message. Messages can be triggered in the following scenarios:

- The OBD recognizes a critical error within the vehicle

Figure 3.8: Prompt to form a corridor (based on[1])

- The OBD recognizes a set off emergency system (e. g. airbag) or an abrupt lane change or braking maneuver.

The message is then propagated in direction of the vehicle's origin and is depending on the gravity of the trigger (e. g. crashes are weighted higher than rapid lane changes).

3.2.2.4 Risks and conclusion

The proposed concept of an awareness enhancing Driver Workload Management Systems shows how to support vehicle drivers with VANET based technologies in situations, in which traditional Driver Workload Management Systems fail due to their limited connectivity and range. The system's feasibility appears plausible; especially the idea of warnings about emergency vehicles is widely supported among researchers[1]. However, the actual usefulness of such a system depends on the adaption of VANET among the population, as propagation of VANET messages only works with a relatively high sender density (10% equipment rate minimum([1], VI).

3.3 Socio-economic aspects of VANET

In this part of the paper we will talk about the different challenges that have to be mastered to get a VANET-system into our society and some of the controversy different approaches bring with them.

3.3.1 Successful market introduction

For consumer technologies like VANET there are two mechanisms that will lead to a successful market introduction: The first one is that there has to be a visible added value of the technology for the end customer, while the second one says that a market introduction of a product is always successful if there is a regulative order that does not leave any alternative technologies[15].

3.3.1.1 Added value for customer

We focus on the added value generated by implementing the previously mentioned warning systems to enhance the awareness of motorists. The main feature of this application

is to improve road security, according to the Sartre (Social Attitudes to Road Traffic Risk in Europe) Report 82.6% of European drivers were concerned about Road safety[16] Although this number has decreased by 3.3% since the survey in 2002 it is still the biggest factor in deciding which car is bought. This shows the number of people that actually care about the safety degree of their car. The number people that care about VANET is even bigger, as equipping a car with a VANET-system not only increases safety for its passengers but also other road users like pedestrians. This is done by helping a driver to be more aware of his surrounding, telling him upfront when he will have to break. The second most important factor when buying a car is also addressed, reduced fuel consumption. Everybody who has seen a car commercial in the last ten years knows that the consumption index (liters/100 kilometer or miles/gallon) of a new vehicle is always mentioned. Especially since the oil price skyrocketed in 2007/2008 fuel efficacy has become very important to most of the drivers. On the long run driving a more expensive car with a lower fuel consumption will probably be cheaper than a cheaper car with higher fuel consuption. The emergency vehicle warning system brings two benefits for the costumers to the table. The first one is that if such a product would be used by the majority of the population, the average rescue time in an emergency situation would be drastically shortened. The question that presents itself here is whether this would be an incentive to pay more for a car or not. Because this is the first factor that is not directly beneficial to the operator or passengers of the car it requires a more altruistic way of thinking and this leads us to think that the integration of such a system would not be seen as a big added value. There is a precedence case of that, the Swedish car producer, Volvo introduced a pedestrian airbag system in its 2014 Volvo V40. They claimed that the chance of a pedestrian surviving a crash with moderate speed (20-50km/h) would be increased. But even though this system was added without an increase of the price the sales number of that particular model did increase while the rest of the brand grew significantly[17]. That fact strengthens our opinion of an altruistic motive not being a driving force in the decision of buying a new car. The second benefit only applies to the relatively small customer group of emergency vehicle owners/drivers. The benefits of comfort and safety are the biggest in this case. And we would imagine that every new vehicle bought would be fitted with a VANET-system, probably vehicles in use would, if possible be retrofitted with the technology. Another benefit customers would gain by buying a car with a VANET-system would be the the "Driver Workload Management Systems" that make driving a less stressful and more comfortable experience. If the driver can rely on a system and can let it take control over little tasks such as maintaining the distance to the car driving in front or even a small thing as warning him of a oncoming emergency vehicle when he wants to listen to louder music, the experience of traveling is a lot better.

3.3.1.2 Regulative order

In Switzerland alone the costs of road crashes per year are at 12.7 billion Swiss Francs[18]. Those costs are carried at least in part by the public. Therefore, decreasing the number of accidents on the road is valuable for the tax money. This is a reason for the government to create laws which demand the use of a VANET system cars. There have been regulations e.g., seat belts and Anti-lock braking system (ABS) are mandatory in every new purchased car. In oarder for a regulative order to be issued, the effectiveness of the technology has to be proven first. With the mentioned technologies this is not a problem, they can easily be tested in crash tests and the benefits will show themselves after slow introduction to the market. Here is where the biggest problem of VANET comes into play. Before the effects of its use in traffic can be shown, there has to be a certain penetration of the market.

3.3.1.3 Market penetration

Studies show that the system will need 10% of the cars using VANET to make the network usable and the system starts to show its potential. Assuming that 50% of all newly produced cars are equipped with a VANET-system, it would take approximately 3 years to get to the 10%[15]. As the network effect of this technology is really strong, no car manufacturer is going to increase sales by implementing a system alone. The common strategy to develop new technology that is used by manufacturer is equipping the luxury cars first and see customers react will not work either because the time needed to reach the 10% penetration is too long.No benefits will be seen by customers, so they probably wouldn't pay extra anymore. Making a VANET-system available as an aftermarket product might be difficult as well, as the system needs access to sensor data and the system might be limited due to the fact that it is not integrated to the bus system of the car.

1. Every newly built car is equipped with a basic VANET-system, that can generate and forward messages with all the needed data. There is no additional cost for those units that the customer has to pay for. This step is only to increase the penetration rate as fast as possible. The vehicles only communicate with other vehicles or roadside units. At this stage, there is no communication to the driver and the system is not marketed.

2. The buyers are offered several options to upgrade their cars with interfaces that allow them to profit from communication with RSUs, such as updated maps or multimedia streaming (see 4.). The sale of those upgrades finance themselves as well as partially finance the basic unit installed in every car.

3. Another part of the initial costs is carried by the manufacturer. As he benefits from all the installed systems and will be able to exploit the functionality of inter car communication later on (see 5.).

4. Insurance companies or the government might want to invest in the refinancing of the units by offering their costumers lower insurance fees / lowering taxes if they choose certain options.

5. Once the 10% penetration is reached, the human machine interface (HMI) for car to car communications can be bought as an upgrade for new cars. The possibility of activating this feature in a vehicle with the basic unit should be given, some of the initial investment might be regained (see 3.).

3.3.2 Controversy of new safety technologies

Due to new safety features in motor vehicles the deaths in road accidents have overall decreased. But there is a inequality of the development between different socio economic groups.

> "Although overall MVA death rates declined during this period, socioeconomic differences in MVA mortality have persisted or worsened over time.We found larger mortality decreases among the more highly educated and some evidence of mortality increases among the least educated." [22]

The fact that the mortality of less educated drivers compared to the mortality of higher educated motorists increased during the 15 year period of the research (1995-2010) was not expected. With the introduction of a VANET-system into new cars this difference would certainly increase even more. More expensive cars have the better safety features

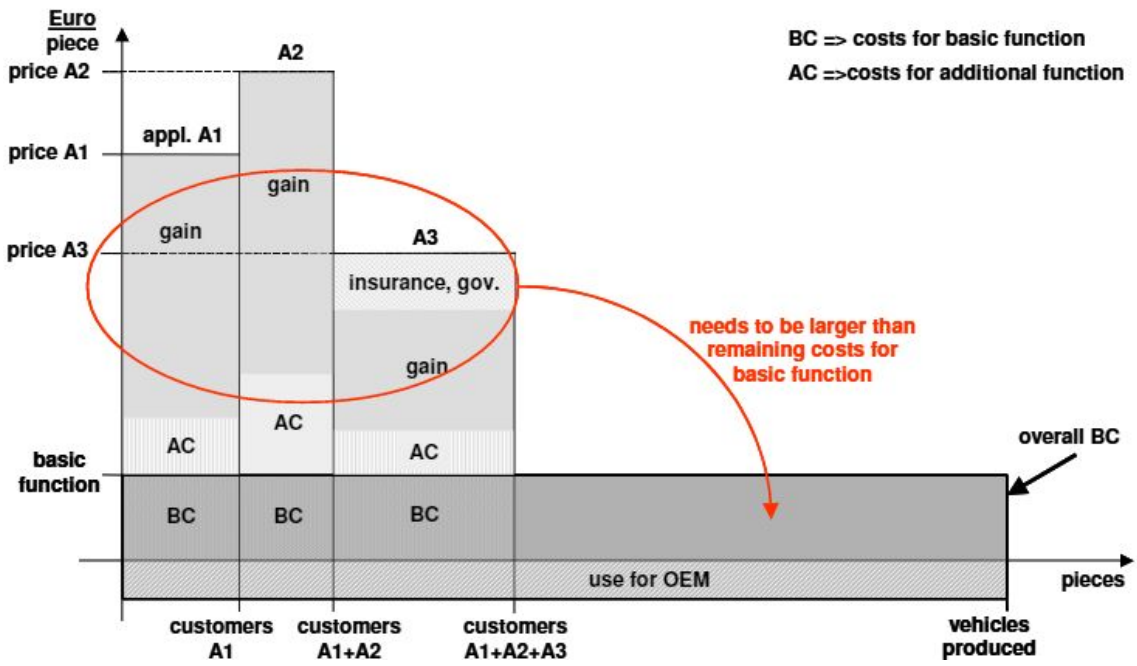


Figure 3.9: Market Introduction strategy visualisation (based on[15])

and new cars are safer but is that really necessary with the VANET-system? By creating a unit that could be retrofitted in some older cars would make the system work against these socioeconomic inequalities. The financing of the development and implementation of those units could also in part be carried by insurance companies, the government and the manufacturers who all stand to profit of a wider spread system and a bigger penetration rate.

3.4 Potential customer reception

In order to indicate, whether a new technologie is going to be successful when introduced to the market and whether customers start to think about buying the options, we have to talk about the needs and benefits of the different applications. Additionally the alternatives that exist and could possibly render the new technology obsolete will be mentioned.

3.4.1 Safety applications

The vast majority of the population values new safety features and wants to have a car that is safer than the old one[16]. Every vehicle has the capacity to broadcast the relevant data gathered by its internal sensors and can also receive this data from other vehicles and interpret it.Because of the direct connections between the vehicles the latency of this way of communication is very low, which is critical in a safety related environment where a few milliseconds matter.

3.4.1.1 Real Time traffic data

The vehicle has the capacity to gather real time traffic data by communicating with other cars and using the internal sensors. As an example, the vehicle realises when an other car is in it's blind spot when the driver is trying to shift the lane on the motorway. It then warns the driver to not change lane. To use a system such as this, the information shared via VANET would have to include location and speed of the cars. Such a system has a great value for the customer because not only can it prevent accidents from happening

but also makes riding a more comfortable experience, you don't have to worry all the time about a car you might overlook. But even though the benefits of such a system are great, they do not really matter because several big car manufacturers already use a system similar to this without having to resort to VANET, they use a small radar to detect cars that cannot be seen in the rearview mirror and most of them indicate it by flashing a light on the mirror so you instantly see it.

3.4.1.2 Road hazards

Using transferred sensor data from other vehicles the car can anticipate and avoid an eventually dangerous situation. For instance, a car driving in front of our car recognises ice (or oil etc) on the road due to traction loss on one of its wheels. It can now broadcast this information to all cars coming behind it, which can inturn then inform their drivers to take action. As of yet there is no other technologie on the market that could communicate or find and avoid the location of ice on a road. But modern cars are equipped with very sophisticated traction control systems that in most cases could deal with ice on the road. Other hazards such as debris or potholes can already be detected by several newer upper class car models, using radar systems.

3.4.1.3 Warning transmissions

Every motorist knows the problem that you want to warn other road user about something. This might be that the somebody forgot to turn on his lights or that there is a police radar station ahead. In some instances these warnings can be essential. An example where the booth of a car opened by itself or that an irregular amount of smoke is coming from underneath another car, those are safety relevant potentially very important messages that would be really hard to detect by a machine, which is probably the reason why there is , as of this moment, no such technology in use.

3.4.2 Non Safety applications

The benefits of having a VANET are not that the Non Safety applications of VANET can be separated into two different categories, commercially used applications and applications for convenience. Using VANET on these application one of its biggest advantages, the low latency does not matter anymore.

3.4.2.1 Commercial application

Whereas most of the data for other applications come from other cars all commercial applications of VANET need a third party involved that provides a service or data sets required.

Location based services

These are probably already known by the name Point Of Interest (POI) commonly used on navigation devices and smartphones. If a user finds himself on the road looking for a restaurant or hotel nearby the system will be able to navigate him/her to the location he/she want to go. This is exactly the same thing any navigation system in a car does perfectly. But by using VANET the hotel or restaurant could expand its marketing by sending some information about table/room availability or if the customer wants it could even send a menu. The POI would have to be fitted with a roadside unit(RSU) so that they are able to control and update by themselves. By using this system one would be able to save unnecessary trips to a restaurant that is already full. this

could also be avoided by calling the restaurant by phone using the handsfree system that is standard in all newer cars, which is a little less convenient.

Internet Access

Granting Internet access to VANET-enabled cars by connecting them to the network via an RSU while using other vehicles as range extenders. The Internet access can be used to download new card material for the navigation system, it could be used to get information about weather conditions or anything else one would want to know. The technology of having a vehicle connected to the internet has already been in use for some years. Premium car manufacturers have connected their cars to the Internet via the cellular mobile phone network. The benefit of a VANET system when compared to the cellular network is, that even if you are in a tunnel or a valley where the network can not reach there could be some RSUs in use. But the same goes the other way around, if there are no other vehicles in the vicinity no one can relay the signal.

Multimedia streaming

The aforementioned application of Internet access needs to be implied, apart from a little exception, for this application. Being able to stream music in real time from the Internet and no longer having to rely on a bad radio signal would be very convenient and make travelling by car less stressful. When driving with children in the back seats they can watch their favourite series on Netflix and leave their parents drive in peace. Improvements in comfort again, that make driving a more pleasant experience. These applications are just an extension of the benefits of having an Internet connection in a vehicle and there is no need to use a VANET-system for them. As mentioned before there is one exception, if it is possible to stream audio and video between cars driving close to each other it would not be a big step to enable a communication between vehicles. using this two vehicle drivers could share messages and talk to each other. The CB radios that are mostly used by truckers would become useless. As mentioned there is technology to talk between vehicles already but it can only be done by buying aftermarket parts and installing a big antenna on your car using VANET would create an open network people can communicate through.

3.4.2.2 Convenience applications

In contrast to the commercial applications of VANET, the convenience applications will have to be supported or even be initialised by the government or the community. These applications can only in some small ways be used for commercial gain and will probably not be economically attractive for businesses because of that.

Parking availability service

The name of this application says it all. A system/service in your car, preferably connected to your navigation system, that tells you where the nearest, free parking spot is available. By setting up RSUs on parking sites that shares information about how many or even which parking spots are vacant, the cars connected to it by VANET could route their drivers to a location with available spots near a given destination. Systems that guide drivers to available parking spots are already in use in bigger cities around the world. Road signs that give a vague direction to a parking sita and a number of how many spots are available can be seen at major intersections. A big advantage of a road

sign based system is that every road user can profit from it, no matter what his chosen vehicle might be. But it has also some major drawbacks compared to an application with VANET. The road signs give no individualised data, which means that you can follow the signs to a parking site but it is not necessarily in the vicinity of your wanted destination. Also the used system has no information about how many cars are on their way to occupy the free spots, so once a motorist gets there he/she certain there will still be a spot for his/her car. In addition to this unreliability there is also the fact that setting up several different, connected road signs costs a lot of money in comparison to the single RSU on every parking site.

Toll collecting service

Several countries use a toll system where road users have to pay for using the highway. Usually there are Toll stations on every highway and drivers are required to stop and pay the fare they are due. Using a VANET application this traffic jam generating method of collecting toll could be greatly reduced. By connecting a RSU to the car collecting the information of every driver coming through the station and billing the fare directly, drivers would not have to stop or slow down, which accelerates travelling and reduces the pollution.

Roadstatus service / Alternate Routes

The general concept of this system is well known because it is already integrated to most navigation systems. Information about road blockages or traffic jams are displayed on the screen and the driver can decide to take an alternate route that shortens the travel time. Using VANET to implement such a system would have the advantage of faster responses to the road status. Every car in the system sends information about its current traveling speed and location. So not only can traffic jams be avoided but also eventual slow moving traffic on a given alternate route can be anticipated and evaded.

Bibliography

[1] Andreas Buchenscheit, Florian Schaub, Frank Kargl, and Michael Weber, *A VANET-based Emergency Vehicle Warning System*, Institute of Media Informatics, Ulm University, Germany

[2] Ewert U.,*Unaufmerksamkeit und Ablenkung, bfu-Faktenblatt 07*, Bern: Beratungsstelle für Unfallverhütung, 2011.

[3] Saurabh Kumar Gaur, S.K.Tyagi, Pushpender Singh, *VANET System for Vehicular Security Applications*, International Journal of Soft Computing and Engineering (IJSCE) ISBN: 2231-2307, Volume-2, Issue-6, January 2013

[4] Ivan Berger, *Standards for Car Talk*, The Institute, March 2007, pages 1-6, `http://theinstitute.ieee.org/ns/quarterly_issues/timar07.pdf`

[5] World Health Organization, *Global Status Report On Road Safety*, WHO Press ISBN 978 92 4 156456 4, 2013.`http://www.who.int`

[6] European Commission - Directorate General for Mobility and Transport, *EU road fatalities*, CARE (EU road accidents database) or national publications, March 2015, `http://ec.europa.eu/transport/road_safety/specialist/statistics/index_en.htm`

[7] Swiss Federal Statistical Office, Transport infrastructure and vehicles, Stock, Level of motorisation, May 2015, `http://www.bfs.admin.ch/bfs/portal/en/index/themen/11/03/blank/02/01/01.html`

[8] Casualties among car occupants, by age and type of occupant, FEDRO, accidents registered by the police, 2013, `http://www.bfu.ch/en/Documents/04_Forschung-und-Statistik/02_Statistik/2014/PDF/E_USV_T_29.pdf`

[9] Adrian Latorre, `http://adrianlatorre.com/projects/pfc/img/vanet_full.jpg`, 2015

[10] J. Luo, J. Hubaux, *A Survey of Inter-Vehicle Communication*, Technical Report, Lausanne, December, 2004

[11] Zeadally, R. Hunt, Y. Chen, A. Irwin, A. Hassan, *Vehicular ad hoc networks (VANETS): status, results, and challenges*, Telecommunication Systems, Vol. 50, Issue 4, August, 2012, pages 217-241

[12] Claudia Campolo, Antonella Molinaro, Riccardo Scopigno, *Vehicular ad hoc Networks Standards, Solutions, and Research* Springer International Publishing, Switzerland 2015

[13] M. Mejri, J. Ben-Othman, M. Hamdi, *Survey on VANET security challenges and possible cryptographic solutions*, Vehicular Communications 1, May, 2014, pages 53-66

[14] P. Green, W. Levison, G. Paelke, and C. Serafin, *Suggested human factors design guidelines for driver information systems*(Tech. Rep.), The University of Michigan, Transportation Research Institute (UMTRI), UMTRI-93-21, 1994.

[15] Patrice Seuwou, Dilip Patel, George Ubakanma, *Vehicular Ad hoc Network Applications and Security: A Study into the Economic and the Legal implications*, 2013

[16] Social Attitudes to Road Traffic Risk in Europe,*SARTRE4 final report*, November 2012

[17] Volvo Group, *Volvo Car Group Retail Sales By Car*, 2015

[18] H. Sommer,O. Brügger, C. Lieb, S. Niemann,*Volkswirtschaftliche Kosten der Nichtberufsunf*Ã¤*lle in der Schweiz: Strassenverkehr, Sport, Haus und Freizeit*,bfu-Report 58. Bern, 2007

[19] A. Aijaz, B. Bochow, F. Dötzer, A. Festag, M.Gerlach, R. Kroh and T. Leinmüller, *"Attacks on Inter Vehicle Communications Systems-and Analysis"*, The Network on Wheels Project, Tech. Rep., 2005. `http://www.network-on-wheels.de/documents.html`

[20] Kirsten Matheus, Rolf Morich, Andreas Luebke,*Economic Background of Car-to-Car Communication*,Carmeq GmbH, Berlin

[21] K.C. Lee, U. Lee and M.Gerla, "Surver of Routing Protocols in Vehicular Ad Hoc Networks," *in Advances in Vehicular Ad-Hoc Networks: Developments and Challlenges*, IGI GLobal, Oct, 2009

[22] Sam Harper, Thomas J. Charters, Erin C. Strumpf, *Trends in Socioeconomic Inequalities in Motor Vehicle Accident Deaths in the United States*, 1995-2010

Chapter 4

Socioeconomic Impact of Missing Parking Availability Information

Adriano Meyer Broyn, Stefan Bublitz, Sacha Uhlmann

Drivers around the globe are having issues finding a parking space on a daily basis. This challenge does not only affect the drivers, who loose time and waste fuel, but it also affects other people and the environment. In the following report the socioeconomic aspects of missing parking availability information is considered, the issue of pricing parking examined, and several approaches that aim to reduce the imapct are discussed.

Contents

4.1 Introduction **75**
4.2 Related Work **75**
4.3 Cruising for Parking **76**
4.3.1 Influence on Traffic of Cruising for Parking 76
4.3.2 Consequences of Missing Parking Availability Information . . . 77
4.3.3 The Effect of Price on Parking Availability 78
4.4 Approaches **79**
4.4.1 ParkU 79
4.4.2 parkITsmart 81
4.4.3 Park & Ride Guidance System in Donostia-San Sebastian . . . 86
4.4.4 SFpark 87
4.4.5 CGE-Model for Parking Policies in Zurich 89
4.5 Conclusion **92**

4.1 Introduction

3650 vehicle kilometers - that is the estimated amount of vehicle kilometers traveled every year per parking space in urban areas [18]. While this number in itself sounds impressive, the consequences of this are more extensive than one might expect at first. Cruising for parking does not only influence the driver by increasing cost for fuel and preventing using the time more productively, but also introduces external effects. It increases traffic, thereby extending the risk of accidents and traffic congestion, and it pollutes the environment. Higher traffic density also affects public transport. It makes it less reliable due to traffic congestions, and therefore there is less incentive to switch to public transportation.
An option to decrease the issues of cruising for parking is to provide drivers with information on parking space availability. With drivers exactly knowing where to park, they can find a free parking space a lot faster and thereby spend less time cruising. This reduces the number of kilometers driven per year and also the side-effects of cruising for parking. This report delivers an overview on the evidence that cruising for parking exists and considers the effects of missing parking availability information. Furthermore, the impact of pricing on parking availability and its consequences are examined. Several existing approaches that aim to reduce cruising for parking, which tackle the issue from different perspectives, are introduced and discussed. Finally, an existing model for evaluating parking policies in Zurich is explored, as such models can serve as a starting point for improving the parking situation in a city.

4.2 Related Work

The issue of cruising for parking is considered in several existing works. A model on how drivers decide if they cruise for parking or choose more expensive parking alternatives is presented in [15]. Drivers cruise more if fuel is cheap, while on-street parking is expensive. Furthermore, results of studies are presented which show that up to 74% of traffic in cities is due to cruising. The negative effects of this, such as air pollution and increased traffic, are explored in [15, 16, 14]. The importance of on-street pricing is also stressed.
Empirical evidence for cruising for parking is also considered by [10] using different models. Conclusions include that there is very little cruising in the Netherlands, but that this is expected as parking prices for on- and off-street parking are the same. Furthermore, the findings include that within-day price differentiation might prove to be an efficient approach to reduce the problem.
The consequences of cruising for parking are explored by [7]. It emphasizes the issue of external effects caused by cruising for parking. In addition, it concludes, that smart parking policies would not only increase societal welfare, but that the negative externalities of cruising for parking are even higher than the ones by traffic congestion in Istanbul.
SFpark, the parking system to manage on-street parking availability and prices in San Francisco, is introduced in [13]. It requires infrastructural changes, but allows for price discrimination depending on demand. SFpark is discussed later in this report. The question if solutions such as SFpark can also be used in other cities is explored in [2]. It examines different options and consequences for Zurich, using a CGE-Model. It is shown that the current situation can be significantly improved, *i.e.* the cruising time reduced drastically.
Another approach is parkITsmart. It uses crowdsourcing to improve parking information availability. parkITsmart includes applications for drivers and parking inspectors, allowing drivers to find parking spaces faster and parking inspectors to work more effectively [5, 6, 9, 18]. ParkITsmart is discussed in greater depth later in this report.

Considering the related work, this report differs in the goal and thereby in the broad and depth regarding the introduction of individual solutions. It provides an overview on the impact of missing parking availability information and available solutions addressing the subject, no new solutions are introduced nor is any new evidence presented. This report should be considered supplementary to the related work and can serve as an entry point towards the subject.
The remaining part of this report is structured as follows: Section 4.3.1 and 4.3.2 provide motivation on why missing parking information and cruising for parking is problematic and should not be neglected. In Section 4.3.3 the problem of setting the right prices for parking is addressed. Then, in Section 4.4 different approaches are considered and discussed.

4.3 Cruising for Parking

Before considering approaches to improve parking availability information, questions arise if missing parking availability information really is problematic and what its impacts are. For this, not only the impact on the traffic in general is considered, but also the effects on the environment and third parties.

4.3.1 Influence on Traffic of Cruising for Parking

The question if cruising for parking has an affect on traffic has been addressed by several researchers. It shows, that the amount of traffic generated by cruising for parking varies widely in different cities [15].
While 45% of traffic on 7th Avenue in Brooklyn, New York, is caused by vehicles looking for a parking space, it is very low in other cities and countries [7]. For example, the average time for cruising for parking in the Netherlands is only 36 seconds per car trip [10]. However, it is not an isolated problem for Brooklyn. In SoHo, Manhattan, New York, 28% of traffic is due to vehicles cruising for a parking space. The cruising for parking problem does not only exist in New York. 16 studies between 1927 and 2001 (Figure 4.1) have shown that the average share of traffic due to cruising is 30% and that vehicles on average search 8.1 minutes for a parking space [15].

YEAR	CITY	SHARE OF TRAFFIC CRUISING	AVERAGE SEARCH TIME (minutes)
1927	Detroit (1)	19%	—
1927	Detroit (2)	34%	—
1933	Washington	—	8.0
1960	New Haven	17%	—
1965	London (1)	—	6.1
1965	London (2)	—	3.5
1965	London (3)	—	3.6
1977	Freiburg	74%	6.0
1984	Jerusalem	—	9.0
1985	Cambridge	30%	11.5
1993	Cape Town	—	12.2
1993	New York (1)	8%	7.9
1993	New York (2)	—	10.2
1993	New York (3)	—	13.9
1997	San Francisco	—	6.5
2001	Sydney	—	6.5
	AVERAGE	30%	8.1

Figure 4.1: Overview of Study Results [16]

Even though the numbers vary greatly in different cities, it is clear that cruising for parking exists and must be addressed. Especially since the vehicles cruising for parking increase the already high traffic, making "a bad situation even worse" [18].

4.3.2 Consequences of Missing Parking Availability Information

A large portion of traffic in urban areas is due to cruising for parking, and several stakeholders are affected by this. First of all, cruising for parking negatively influences the drivers themselves. A person visiting an urban area once per day for a year, using the average cruising time from the previous section, will spend nearly 50 hours just cruising for a parking space. Furthermore, drivers will burn fuel required for driving an extra 50 hours. Cruising for parking in Westwood Village, a district in Los Angeles, with 470 metered parking spaces, was analyzed by [14]. The total search time, *i.e.* the time spent on cruising for parking, in Westwood is estimated to be 426 hours per day. That results in a cruising time of 54 minutes for every parking space, every day [14]. Furthermore, the distance traveled by vehicles cruising for parking in Westwood is "equivalent to 28 trips around the earth or two round trips to the moon" [14].
Moreover, cruising for parking does not only affect the drivers themselves. While cruising for parking in Westwood wastes 11 years of drivers' time every year, it is also responsible for 178'000 liters (47'000 gallons) of fuel consumption and pollutes the environment with

728 tons CO_2 every year. Drivers looking for parking spaces are also increasing the risks of accidents, rising the possibility of traffic congestion, slowing down traffic and contributing to noise pollution [14, 7].
These numbers are only estimations and it is difficult to generalize them (also see [7]. However, they still demonstrate that cruising for parking is a serious issue that must be addressed.

4.3.3 The Effect of Price on Parking Availability

The time spent cruising for parking varies greatly over many cities. Before exploring the underlying reasons for this characteristic, the difference between on- and off-street parking must be considered. On-street parking is often cheaper than off-street parking and located at more convenient locations [17, 15].
One of the main factors influencing cruising time is the cost of on-street parking [10, 7]. In locations where on-street parking is similar expensive as off-street parking, the cruising time is often a lot lower. However, the decision if one parks on-street or off-street is not only influenced by the price. [15] suggests that the decision for cruising is influenced by several factors:

- Price of on-street parking ($/h)
- Price of off-street parking ($/h)
- Parking duration (h)
- Time spent searching for on-street parking (h)
- Fuel cost of cruising ($/h)
- Number of people in the car (individuals)
- Value of time spent cruising ($/h/person)

As the formula shows, the decision is affected by a multitude of factors. Some of these factors are based on assumptions. For example, the expected time spent for searching can vary greatly from the actual time looking for a parking space. Other factors, such as the value of the driver's own time and the time of other passengers in the car are also difficult to quantify. In addition, it is also questionable if people are aware of the price for fuel while cruising for parking. However, considering the formula, it is clear that prices for the different parking options influence the decision and, even though they are not the only factors, can be used to change the behavior of drivers.
According to [15], underpriced parking is one of the main reasons why drivers cruise for a parking space. Increasing the prices can decrease the time for cruising substantially. Increasing the price for parking thereby also reduces the risk of peak-hour congestion [4]. This is further supported by the situation in northern European countries, where a lot less time is spent on cruising for parking, as off-street parking and on-street parking are offered at the same prices [10].
Figure 4.2 visualizes the situation in Westwood. Underpriced on-street parking leads to a situation where all parking spots are occupied and people need to cruise for parking. However, when the prices are increased, the occupancy rate decreases and drivers will find a parking space immediately.
Defining the right prices for parking is not a trivial task. [16] suggests, that only trial-and-error will reveal the right price. However, statically increasing the prices for off-street parking reduces many risks. Too high prices however are also risky: off-street parking "spaces remain empty, nearby merchants lose potential customers, workers lose jobs, and

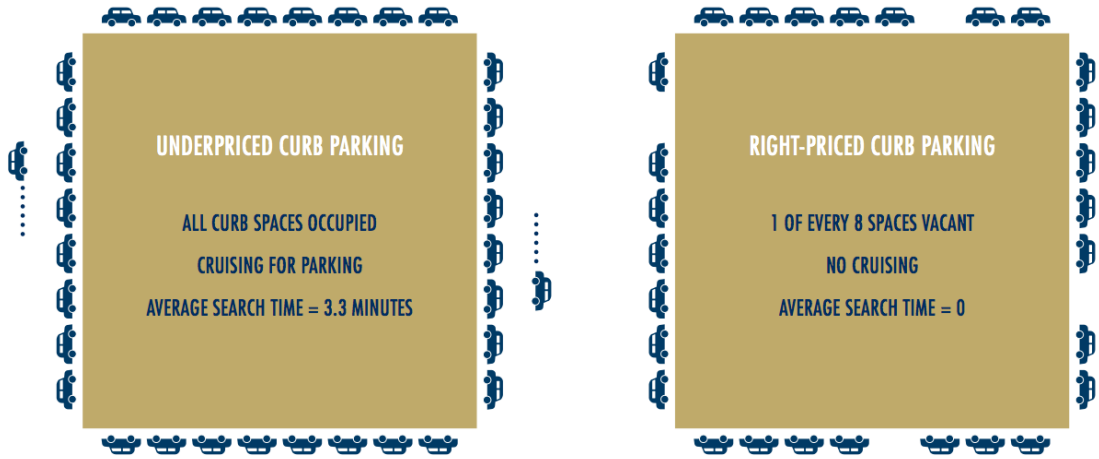


Figure 4.2: On-Street Parking Prices and the Impact on Cruising [16]

cities lose tax revenue" [13]. In cities where citizens are used to free or cheap on-street parking, introducing or increasing the rates might be politically dangerous. Furthermore, it must be discussed if parking spaces should be considered as common good or of it is acceptable to have siltations where more wealthy people can park at more convenient places (on public ground) than others.
So, while increasing the parking prices might not be a solution without risks, price adjustments in general should not be neglected and can in fact reduce traffic. However, statically increasing all on-street prices is not ideal, and alternatives must also be considered.

4.4 Approaches

Several existing approaches to reduce the cruising for parking issue are introduced and discussed. The different approaches vary greatly regarding costs for implementation, required infrastructural changes and time required. Furthermore, the entity that must become active differs in several approaches. As the complexity and consequences of the introduced approaches varies, some approaches are discussed in greater length than others. The list of introduced solutions in this report is not exhaustive, several other approaches exist. However, it is the opinion of the authors that these solutions provide an overview of the different perspectives from which the subject can be addressed. Some solutions require governmental spending, others tackle the issue with information crowdsourcing and even others introduce a new layer to the existing market. The first approach considered is rather a market and not a technical innovation.

4.4.1 ParkU

ParkU is a private company that offers parking spaces in over 15 cities in Austria, Germany, the Netherlands and Switzerland. There are currently over 5000 parking spaces available via ParkU. The parking spaces offered by ParkU are not owned by it, ParkU acts as an agency between private entities that have parking spaces available and drivers looking for a parking space. The different entities and how they are related to each other are visualized in Figure 4.3.
Drivers looking for a parking space sign up via website or mobile app and reserve a parking spot. Parking spaces can be reserved up to 30 days in advance and for time spaces of 30 minutes or longer. The drivers pay for ParkU via the mobile app, using Paypal or Credit Cards [12]. Using ParkU, drivers can reduce their cruising time, as they exactly know

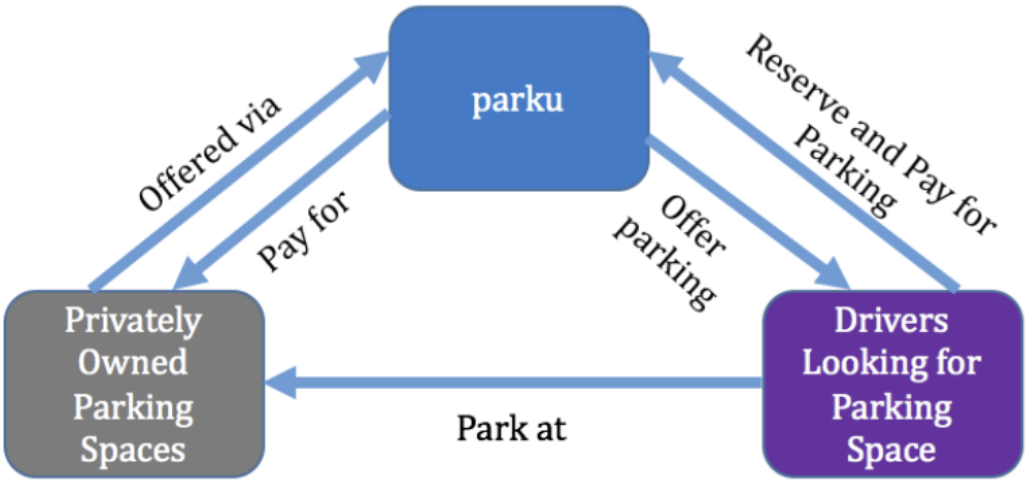


Figure 4.3: ParkU: Entities and their Interactions

where to park. It is theoretically possible that a parking space is occupied (illegally) by another driver [18]. In such a case, ParkU will provide drivers immediately with another parking space and drivers receive 5 hours of free parking. Regarding ParkU, illegally occupied parking spaces have not been an major issue [3].
The approach by ParkU has several advantages and disadvantages. ParkU is easy to understand, does not require any changes to the existing infrastructure nor governmental action. The approach can be transferred to other locations relatively easily. However, since it does not affect any existing infrastructure, it also has major limitations. ParkU does not show the parking availability of other parking spaces, therefore drivers do not know in advance if the ParkU parking space will really be the best available solution to them. Furthermore, it is questionable if a solution such as ParkU can scale endlessly. If too many people use the solution, it is likely that the availability of parking spaces will be rather low. As ParkU is market-driven and the prices are not regulated by the government, issues discussed in Section 4.3.3 can occur. However, as ParkU mainly offers parking spaces which were not offered before, it is unlikely that people will pay much more than for different off-street parking.
The main advantages include that parking space availability is provided and that a parking space can be reserved. It allows people to monetize a parking space which they own but do not need all the time. Therefore, ParkU does not only provide parking availability information, but also increases the number of available parking spaces. As they are privately owned, no governmental spendings are required.
It should be mentioned that ParkU is not the only organization that offering such a solution. Table 4.1 provides an overview of different vendors.

Vendor	Parking Spaces in
ParkU	Austria, Germany, the Netherlands, Switzerland
Parking Panda	United States
JustPark	Worldwide

Table 4.1: Overview Parking Agencies

Some of them operate internationally, such as JustPark [8] and Parking Panda [11]. JustPark, originally called ParkatmyHouse, also allows for the reservation of parking spaces

in public places, such as schools, hotels and car parks. Parking Panda differs as the offered parking spaces are usually in existing parking lots and garages, but also allows for searching and reserving parking spaces via app and website. As the general idea is always similar, these will not be discussed in this report and the focus will be on approaches from different perspectives.

4.4.2 parkITsmart

parkITsmart is a project implemented by students at the University of Zurich. It aims to provide parking availability information for drivers in Zurich in order to reduce the time required for finding a parking place. Secondary goals achieved by such a progress are time savings gained by the driver, decrease in generation of avoidable traffic load and less generation of CO2 emissions. At the time of writing, parkITsmart includes the main application called Parking Monitoring and Management System (PMMS), as well as three enhancements. The mentioned improvements include the Specific Activity Monitoring System (SAMS), the Parking Space Marketing System (PSMS) and lastly the Parking Availability Prediction System (PAPS). Hereafter, the various applications are being explained and analyzed.

The Parking Monitoring and Management System

The PMMS presents the foundation for the parkITsmart project. It offers two different applications to ensure a quality - oriented information service. One of these applications is a platform - independent mobile program, which can be utilized by drivers to receive news about free parking spaces in their surrounding area. The other one is an Android application which can be used by parking inspectors to both increase their working performance and provide the occupancy status for parking spaces to be used by parkITsmart. The architecture described including the elements in the background are depicted in Figure 4.4. It shows, that both applications utilize and maintain data from the same database.

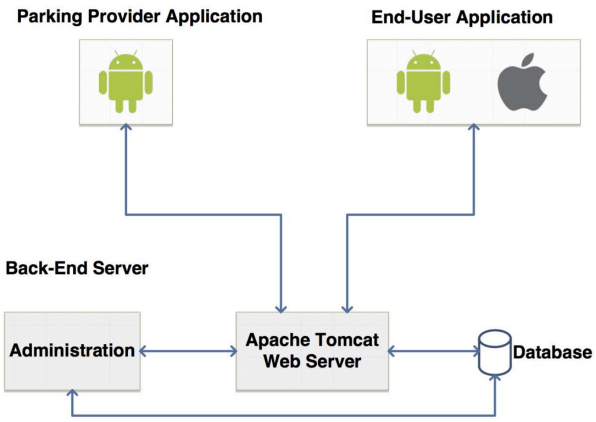


Figure 4.4: Basic Architecture of the PMMS [18]

The end-user application, intended for the use by the drivers, works as follows. The consumer wishing to find a parking place can request the program to find a parking spot for him. Upon deciding on leaving the car on a specific spot, the driver can manually inform the application about his final decision. This helps PMMS to keep occupancy information as relevant as possible. Besides these basic functions, the client application offers advice about on-site regulations and a guidance system routing the user back to his car in order to leave the parking position.

The parking provider Android application is intended to help parking controllers checking cars for valid permits. This support is provided by integrating a Near Field Communication (NFC) tag and a Quick Response (QR) code to the existing parking permit. This feature grants the parking inspector the ability to scan the QR code of the parking license using his mobile phone and check, whether the permit is valid. At the same time, PMMS updates the occupancy status of the parking spot nearest to the NFC tag within the permit as rented for further consideration upon occupancy requests.
Two separate evaluations were carried out to test whether this basic version of parkITsmart incorporating PMMS has a positive effect both on drivers and parking inspectors.
To assess the routing performance of drivers working with parkITsmart a model capturing the basic searching for a parking spot behavior has been developed. This design contains a grid established by 10 x 10 squares, where each square represents 100 meters. Moreover, each square is in either one of two states, *free* or *occupied*, informing about the occupancy status of the parking spot within its boundaries. Within this grid, separate vehicles, starting from random points, had the task to find a parking spot and reach a randomly selected destination. Since moving by car is faster than walking, localizing a parking spot near the destination is an important aspect. With respect to routing behavior, random routing is compared with routing using information provided by parkITsmart. Running this simulation 100 times showed, that using parkITsmart significantly reduced the time required to reach the destination.
The achieved efficiency improvement for parking inspectors has been tested using the following mathematical model.
The goal of this evaluation is to compare the time required for a parking inspector to check a certain number of vehicles utilizing services provided by parkITsmart to the normal case when having to check permits manually. Regarding the parking inspector's characteristics, let t_w denote his walking pace. Moreover, express the time required to check one car with t_c and lastly the time needed to prepare a fine with t_f, both in seconds. With respect to the environment, N describes the number of inspected vehicles and I the number of issued fines. Together, the total time required to check vehicles and generate fines can be expressed with the following formula:

$$t_{total} = t_w + (t_c \cdot N) + (I \cdot t_f) \tag{4.1}$$

The evaluation compares the parking inspector's workflow when having to check a total of $N = 20$ cars. The average walking pace t_w has been found to be $4\frac{\text{km}}{\text{h}}$ on average. While manually checking a parking permit for its validity needs 5 seconds, taking advantage of parkITsmart to scan the QR code reduced the required time to 3 seconds. Finally, preparing a fine manually required 120 seconds, while parkITsmart completes this task within 30 seconds. These properties lead to the following total times in seconds required to check 20 vehicles:

$$\begin{aligned} t_{\text{total}}^{\text{WithoutApplication}} &= 90 + (5 \cdot 20) + (120 \cdot 2) \\ &= 430 \end{aligned} \qquad \begin{aligned} t_{\text{total}}^{\text{WithApplication}} &= 90 + (3 \cdot 20) + (30 \cdot 2) \\ &= 210 \end{aligned}$$

Comparing both times, it becomes clear that due to the saving of nearly half of controlling time, using parkITsmart offers a significant efficiency increase compared to the manual process of checking permits.

The Specific Activity Monitoring System

SAMS is an extension to the existing PMMS aiming to utilize additional information provided by the vehicle and its current surroundings. Figure 4.5 depicts the change in architecture by introducing SAMS. The additional element inside the end-user application

indicates a new source of information communicating with the web server. Hereafter, the motivation for putting this module into practice, its implementation and an evaluation are described.

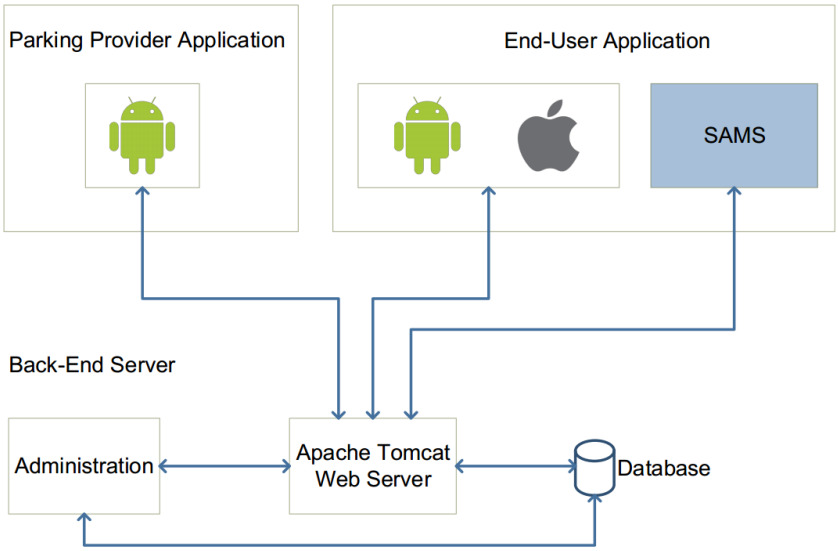


Figure 4.5: Architectural Changes induced by the Specific Activity Monitoring System [9]

The PMMS requires the user to actively specify, where the car is parked for the application to be able to update the occupancy status of the parking spot. In order to automate this process, the application requires additional information.
The implementation of SAMS includes a feature called iBeacon, which allows an application to estimate the distance between the mobile phone its running on and the iBeacon enabled device. Moreover, those iBeacon devices can send messages to the smartphone when entering its signal range, enabling the application to include additional information sources. SAMS mainly makes use of the aforementioned feature, while PSMS utilizes the latter one.
In order for SAMS to make use of iBeacons, the user has to place a device within his car which is able to send iBeacons to the parkITsmart application. The reference implementation chose Gelo Beacons to handle the sending of iBeacons to the program. Since this feature only allows the application to measure its distance to the Gelo Beacon, additional sources providing information about the state of the vehicle can improve the performance of SAMS. Therefore, SAMS introduces a second source of information directly connected to the car. The optional OBD-2 device provides current data regarding the vehicle's condition, *e.g.* the engine speed in revolutions per minute or the vehicles speed in kilometers per hour. This helps SAMS to identify whether the car is currently parked.
In order to evaluate SAMS in terms of additional features compared to PMMS, SAMS was used to detect parking behavior and tested in terms of energy efficiency, parking events accuracy and lastly costs. While in terms of energy efficiency, power consumption of SAMS seems to be quite high, the additional features offered by its implementation are making up for its power utilization. With regard to the parking events' accuracy, false positives and negatives were calculated as follows:

$$\text{False}_{\text{Positives}} = \frac{\#\text{Wrong Detected Events}}{\#\text{Actual Events}} = \frac{1}{18} = 0.06 \qquad \text{False}_{\text{Negatives}} = \frac{\#\text{Not Detected Events}}{\#\text{Actual Events}} = 0$$

While the false negatives result shows that all events that happened where recognized,

false positives draws another picture. The wrong detected event happened when the vehicle was standing due to a traffic jam or traffic light. Within the cost evaluation, different implementations of SAMS were being compared in terms of their costs and features. The lowest cost results in using the combination of smart phones, iBeacon and OBD-2.

The Parking Space Marking System

Another extension, the Parking Space Marking System, offers users of parkITsmart the possibility to provide meta information about parking spots. This data includes the exact location, orientation, size and regulations necessary for parkITsmart to present valid occupancy information about a parking location. In terms of architecture, it is integrated in the parkITsmart environment, meaning that it uses the same database as PMMS and the other extensions. The complete architecture can be seen in Figure 4.6.

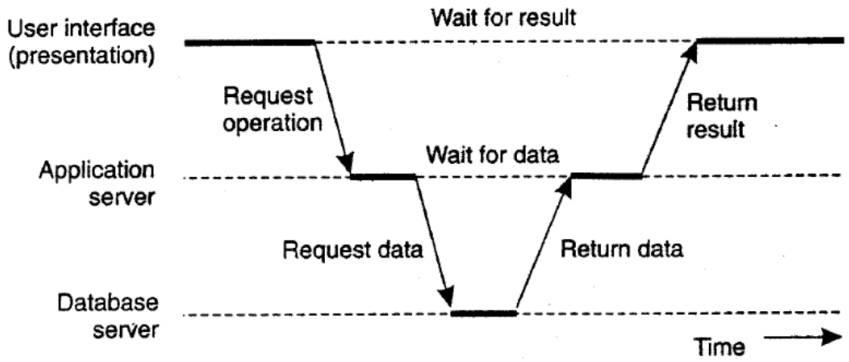


Figure 4.6: Architecture of the Parking Space Marking System [6]

PSMS needs to differentiate between parking inspectors and end-users. While including both parties increases the likelihood of receiving meta data about parking spots, the level of trust towards both parties differs. Although parking inspectors are credible users and therefore assumed to only provide valid meta information, end-users might exhibit malicious behavior to limit the usefulness of parkITsmart. This behavior can manifest in intentionally providing faulty data and therefore has to be validated in order to prevent storing it into the database. Hereafter, the actual process of adding a parking spot including its meta information is described. Moreover, the additional features offered for the parking inspectors and an evaluation analyzing PSMS are discussed.

In order to add new parking spots to parkITsmart, a user first has to define a parking area, which will include a group of parking spaces. All encompassed spots by a single area have common characteristics in terms of regulations, *i.e.* prices and times, orientation, *i.e.* diagonal, straight or aside, and size. Upon creating a parking area, a user can mark, adjust or delete parking spaces within such an area by specifying a single point. The calculation in terms of covered ground of the parking spots within the area is performed by the server based on the placement of the points and the provided parking area meta information.

Provided information by end-users needs to be validated in order to get marked as valid and stored in the database. This validation process includes the requirement, that each proposed parking spot needs to be authenticated by either a single parking inspector or two other end-users. This verification process allows users to mark a suggested parking spot as valid, partially valid or invalid. As soon as one of the two end-users does not mark a proposed parking spot as valid, and therefore as partially valid or invalid, a parking inspector is required to specify whether it is a valid parking spot or not.

As an additional feature for parking inspectors, PSMS utilizes the iBeacon component introduced by SAMS. Vehicles equipped with an iBeacon can be remotely asked if their

parking permit has already been checked. This reduces the amount of walking a parking inspector has to perform to check vehicles.
The evaluation includes two separate assessments, one for the parking marking system and one of the parking inspector application extension. While the parking marking system has been evaluation in terms of usability and time efficiency, the parking inspector feature assessment includes among other areas the time efficiency.
The usability evaluation revealed the necessity for a multilingual support and some UI improvements regarding the map and button behavior within the application. On the other hand, time efficiency analyzed the time required to digitize parking spaces while utilizing PSMS. It concluded with requiring approximately 27 minutes to gather information about 10 parking areas and needing additional 15 minutes to digitize the gathered data using PSMS.
The time efficiency evaluation of the parking inspector's application is an extension to the mathematical model PMMS used to assess its basic version. Including PSMS, the time required to check a parking permit for its validity is further reduced to $t_c = 2$ seconds, leading to the following total time in seconds:

$$\begin{aligned} t_{\text{iBeacon}} &= 90 + (2 \cdot 20) + (30 \cdot 2) \\ &= 190 \end{aligned}$$

Comparing this case to both manually checking and utilizing only PMMS, it turns out, that PSMS offers an additional increase in efficiency.

The Parking Availability Prediction System

Finally, PAPS is the last extension of PMMS. It combines several sources of data to reliably predict the occupancy status of parking spots in real-time and near future. Besides these prediction mechanisms, the comprehensible visualization of the provided information is an important aspect. Figure 4.7 depicts the various sources of information, including their dependencies, which are taken into consideration when calculating predictions.

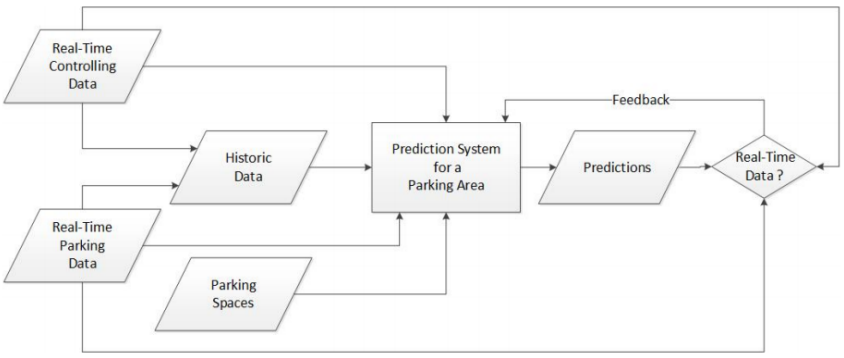


Figure 4.7: Data Sources within the Parking Availability Prediction System [5]

As a prediction function, PAPS uses the Scaled Factor Method, which works as follows: For a given parking area, the number of available parking spaces is equal to the total number of parking spaces it encompasses. For each incoming parking event within that area, the availability is reduced by $\frac{1-fp}{p \cdot (1-fn)}$, where fn identifies the false negative, and fp the false positives rate. Moreover, p describes the percentage of drivers using PAPS or SAMS. On the other hand, leaving a parking spot within that area increases the availability by the same rate. The prediction is then being made by comparing the availability rates of multiple parking areas.
The evaluation of this process calculates the Root Mean Squared Error (RMSE), which is defined as follows:

$$\text{RMSE} = \sqrt{\frac{\sum(\text{Predicted} - \text{Real})^2}{N}}$$

Here, *Predicted* identifies the number of predicted free parking spaces within an area and *Real* the actual number of free spaces. N is the number of predictions. The evaluation shows that the prediction system is more accurate with increasing penetration rate, *i.e.* if more drivers are using SAMS or PAPS.

4.4.3 Park & Ride Guidance System in Donostia-San Sebastian

Civitas Archimedes is a project co - financed by the European Union. It encompasses various projects in different cities around Europe and aims to revolutionize urban traffic in order to achieve significant improvements in topics around energy consumption and environmental sustainability. Subsequently, the project "Park & Ride Guidance system" located in Donostia-San Sebastian is going to be explained and analyzed.

The city council of Donostia-San Sebastian follows a sustainable urban transport strategy and wants to use Civitas to expand this approach [1]. The goal of the project is to increase the number of people using environmental friendly transport methods like public transport or bicycle, while reducing the number of cars entering the city. A promising approach to achieve these objectives was the expansion and improvement of an already existing Parking Guidance System (PGS). The currently implemented PGS covers the distribution of signposts indicating the position of different parking locations throughout the city. In addition, some signs imply the occupancy status of the parking location by showing a red or green light.

As an improvement of the previously existing signposts, the use of Variable Message Signs (VMS) is introduced. These signs are computerized panels which can display text on four lines, each containing 15 digits. Furthermore, different colors can be used to highlight important phrases. These VMS panels are placed at the edge of the city on main entry routes and critical points in the heart of the city. They display information about the cities Park & Ride facilities including number of free parking spaces, and an arrow indicating the direction on how to get there. In addition, they are able to display any other information aiming to reduce traffic congestion, like warnings or recommendations.

Alongside those VMS the use of parking availability information signposts is implemented. The location of these signs depend on the placement of the VMS as they are are placed along strategic routes throughout the city. They direct the driver to the parking spaces. Furthermore, the number of free parking spaces within the relevant parking area is indicated.

To handle the amount of data efficiently, the data centralization system was introduced. This system requires real time data from parking area providers about their occupancy status to be able to project this information onto the VMS. This system can also be used to add new parking areas to the system, and specify the information displayed for warnings or recommendations.

The evaluation compares the situation before implementing the project, hereafter referenced as *before*, with the condition after realizing it, hereinafter referred to as *after*. Moreover, the circumstance was included, where instead of VMS only regular sign posts, similar to the already existing ones, were implemented. The latter one was called Business as Usual (BaU). In order to realistically predict the impacts of BaU, an extensive model illustrating many influencing factors to get a result comparable to reality, was developed. The evaluation process is divided into four different areas: Economy, environment, society and transport.

The economy aspect compares various costs necessary to manage the corresponding implementation. The calculation shows that while the fixed costs necessary for putting *after*

into practice are relatively high, the operating costs including the maintenance of the server structure are low. Comparing this to *before*, additional fixed costs of about 180.000 Euro and operating costs of roughly 6.000 Euro per year are required to implement and maintain it. While implementing *BaU* would require an additional 14.000 Euro, little operating costs would incur, since fixed signs do not need major maintenance.
In terms of environment the different emission levels are being analyzed. While both cases, *after* and *BaU*, seem to have a slight increase in emission levels compared to *before* due to increased mobility levels in the city, *after* leads to a slightly minor surge in CO_2 emissions than *BaU*.
In order to evaluate the impact of Civitas on society, surveys before and after the implementation were conducted. They showed that the level of acceptance for the project increased. According to the authors, this development is due to the missing awareness of the participants of the plans of the project. A comparison of *after* with *BaU* is not possible since an actual implementation has not been conducted.
The topic of transport compares the number of cars entering the city by using the permanent automatic traffic counter network placed along the major entry roads. This shows, that while both situations, *after* and *BaU* lead to a decrease in number of arriving vehicles, *after* can significantly reduce the number by 300 cars per day more than *BaU*.
The evaluation showed, that for Donostia San-Sebastian, implementing VMS panels instead of regular signposts was a profitable decision. The proposed goals for Civitas were achieved and several ideas, like the inclusion of on-street parking, are possible extensions for the future.

4.4.4 SFpark

SFpark is a program to adjust curb parking prices in the city of San Francisco. The pilot started 2011 in seven zones of the city. In these zones sensors have been installed to report the occupancy of each curb on every block. In addition the city has put new parking meters that receive the data of the sensors to establish a price according to the occupancy and the time of the day.
With this setup, the aim of this program is to implement a big price reform for on-street parking. The new pricing systems should vary according to time of day and from block to block.
To explain the necessity of such an intrusive interference, the current curb pricing system has to be analyzed. It works as follows: The price is unchanged all day, many cities do not differentiate between locations, *i.e.* they charge the same price everywhere. The consequences of such a price system is a lot of cruising for inexpensive parking space, thus increasing traffic congestion, wasting time and fuel, and polluting the air. On the other side, if on-street parking spaces are overpriced and remain empty, nearby merchants lose potential customers, workers lose jobs and cities lose tax revenue [13].
So the goal must be to find a balance between these two extremes. Nobel Prize economist William Vickrey proposed that curb parking prices should be set "at a level so determined as to keep the amount of parking down sufficiently so that there will almost always be space available for those willing to pay the fee" [19] (via [13]). The application of this statement is to adjust the price so that there will be always one or two open spaces on each block, in other words, the occupancy rate for a block should be not more than 80 percent. This measure would increase the prices in central and touristic areas of the city, but parking spaces a little bit more away of this locations would be cheaper.
The time of the day also plays an important role for the price setting. For these situations parking prices on a weekday at Fisherman's Wharf in May 2012 were evaluated. Before the first changes, the price was $ 3 an hour at all times.

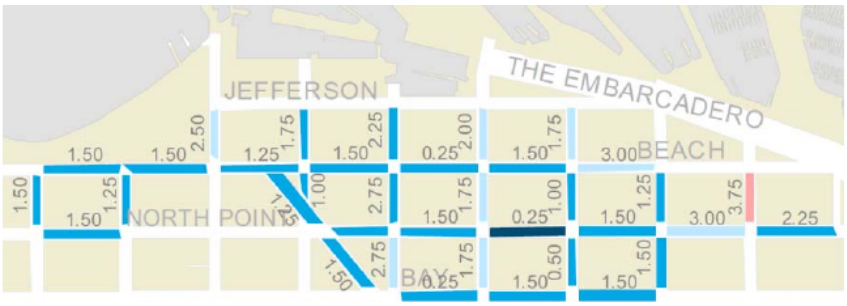


Figure 4.8: SFpark: Parking Prices at Fisherman's Wharf before Noon [13]

Prices were lower than $ 3 in the most streets before noon.

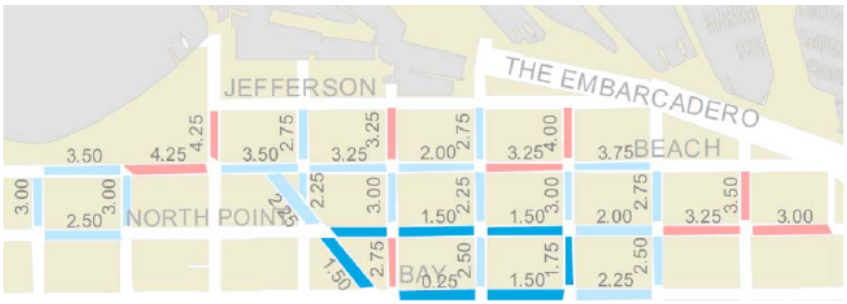


Figure 4.9: SFpark: Parking Prices at Fisherman's Wharf Noon to 3 PM [13]

From noon to 3 pm, in many streets, the prices have increased to over $ 3. However, there are still streets with less expensive parking spaces within one block from the more expensive ones.

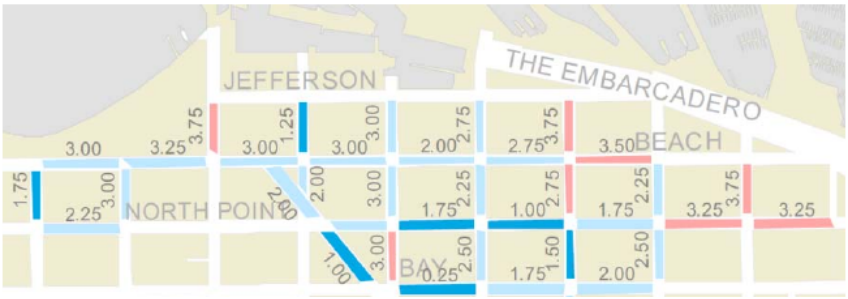


Figure 4.10: SFpark: Parking Prices at Fisherman's Wharf after 3 PM [13]

The prices have decreased a little after 3 pm, but most streets kept their prices over $ 3. Once again, there are still cheap parking places near the expensive ones.
In the end, it can be assumed that every zone has time periods with higher parking demand. During these periods the prices will increase in the critical streets, but there will always be less expensive alternatives nearby.
If higher prices encourage a few parkers to move away from the most crowded blocks, the question arises which people will move first and if they have common characteristics. Long-term parkers would have a greater benefit if they moved to cheaper, but less convenient curb spaces. Also, solo drivers will save more money, if they park a bit further away, because the money that one saves parking in a cheaper parking space is divided by the number of individuals in the car.
These two groups would be the largest ones, but in addition, drivers who arrive early at work could also take the time to park in a cheaper place. This would allow drivers who arrive late to park near their workplace. In the end, there would be lower-income drivers who place a lower value on saving time. This last group would profit more, if they shift their parking locations and walk farther.
These examples show that different prices for different people will create a benefit for all.

Technical View of the Project

In April 2011 the pilot program covered seven zones in San Francisco with 7000 metered curb spaces and 14 public garages. The meters operated daily from 9 am to 6 pm with changing prices at different times of the day, and between weekdays and weekends. The San Francisco Municipal Transportation Agency (SFMTA) established the occupancy rate between 60% and 80%. As a result, the price stays unchanged, if the average occupancy rate falls in this range. Otherwise, the price will rise or fall. There is an upper limit and a lower limit for the prices as well. The minimum price is $ 0.25 and the maximum is $ 6. To set an occupancy rate is very important, to ensure that there will be always one to two free spaces in every block.
At the beginning of the program, the old prices remained unchanged. Two months later, the first prices changes on block level occurred. The prices began to change according to demand. There were six price adjustments in the first year, and at the end results showed that prices increased in 32% of the locations, decreased 31% and remained the same in 37%. On average, the prices decreased during the morning and increased at midday and in the afternoon. In general, the average price fell 1% during the first year of SFpark. In terms of occupancy, there was a considerable progress toward the problem of overcrowding on some block and very low occupancy on other. Table 4.2 shows the results:

Initial Occupancy	**Occupancy after price change**
Below 30%	67%
Above 90%	68%

Table 4.2: SFPark: The Effect of Price Adjustments on Occupancy Rate [13]

Socio-Economic Impact

In terms of fairness, SFpark wants to keep the drivers informed about prices. This is achieved with real-time information via Internet. On the SFpark website, everyone can check a map which shows price and availability of parking in every block for any time of the day. This means, in the future every automobile guidance system can incorporate the pricing data from SFpark. For example, this could be a great benefit for tourists, as they receive real-time information about parking alternatives.
SFpark indirectly also benefits people who do not own a car. Due to the decreased cruising for parking, public transport is less stuck in traffic, and therefore more reliable.
Finally, it must be mentioned that the goal of the program is to optimize the occupancy ratio. The focus is not on maximizing revenue.

4.4.5 CGE-Model for Parking Policies in Zurich

The successful approach to solve the parking problem in San Francisco leads to evaluate the same approach for other cities, for example Zurich. A model to predict the traffic congestion under the impact of different parking policies is needed. This Section builds upon Anne-Kathrin Bodenbender's Master Thesis: A CGE-Model of Parking in Zürich: Implementation and policy test [2].
First, the parking policies in Zurich have to be considered. The prices for off-street parking are very high with disproportional price increases every hour, *i.e.* each additional hour is more expensive than the preceding hour. On-street parking (blue zones) allow free parking for the first 60 minutes using the European parking disc. In some areas, parking for over one hour costs between CHF 1 to 3, between 8 am and 6 pm. On- and off-street parking spaces are always more expensive in the center of the city. To explain the severe parking problem in Zurich, the following fact must be considered: Between 1990 and 2010, the

number of parking spaces has been augmented by 1% (76 additional parking spots), while the number of car has increased drastically more. Therefore, the situation has worsened considerably.
To compare different theoretical approaches to improve the situation, the CGE-Model introduces different scenarios. The basic idea is to observe a simplified network with agents driving through different scenarios. Figure 4.11 shows the simplified network.

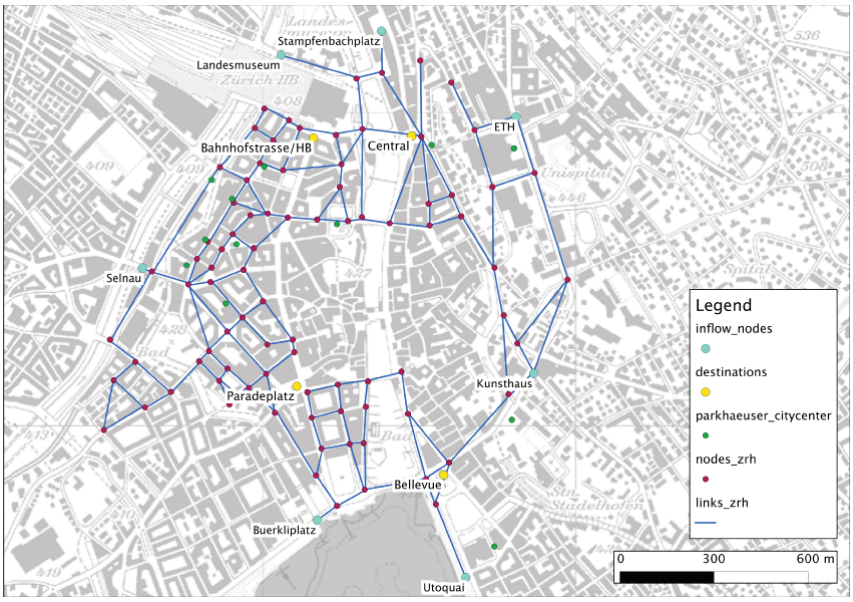


Figure 4.11: CGE-Model: Simplified Street Network of the City Center in Zurich [2]

Agents can enter this network at seven different locations and choose between four destinations in the city center. Every inflow node brings a different amount of traffic to the city. For example every day more cars are coming from Buerkliplatz than from Utoquai. This is called traffic share by inflow node (Table 4.3). Another variable to consider in this model is how much the drivers can afford to pay for a parking space. This depends on the income, so it is assumed that every inflow node is linked to an area of living with higher or lower housing prices and therefore higher or lower income. For example people coming from Utoquai live on the "Goldküste", so it can be assumed that they have a higher income than people coming from Selnau or Stampfenbachplatz. This is called valuation by household (Table 4.4).

Inflow Node	Traffic Share
ETH	19.3%
Bürkliplatz	19.1%
Stampfenbachplatz	16.2%
Landesmuseum	15.7%
Kunsthaus	10.6%
Selnau	10.4%
Utoquai	8.7%

Table 4.3: CGE-Model Zurich: Traffic Share by Inflow Node [2]

Inflow Node	**Valuation by Household (CHF/min)**
Utoquai	2.5
Kunsthaus	2.5
Landesmuseum	2.0
Bürkliplatz	1.5
ETH	1.5
Selnau	1.0
Stampfenbachplatz	0.5

Table 4.4: CGE-Model Zurich: Valuation by Household Type [2]

Every driver who enters the network has a destination. However, many people might share the same destination, that means every destination has a traffic share variable (Table 4.5). Finally, the city garages are marked with a green dot.

Destination	**Traffic share**
Bellevue	33.8%
Paradeplatz	26.0%
Hauptbahnhof	25.3%
Central	14.9%

Table 4.5: CGE-Model Zurich: Traffic Share by Destination [2]

Looking at the different scenarios, a so called benchmark scenario is defined. It reflects today's parking policy in Zurich. Drivers pay a fixed parking fee which is lower for on-street parking than for off-street parking. The average parking time is two hours, and 80% of all on- and off-street parking spaces are being used. The second scenario reflects a similar policy to SFpark. Off-street and on-street parking prices are adjusted to location, time of the day and day of the week. That means, drivers can park at the desired location as long as they are willing to pay the fee.
The third scenario is a policy that ensures on-street parking, called demand-responsive pricing for on-street parking. Garage fees remain identical to the benchmark scenario and the on-street parking prices are adjusted so that there will always be a free spot.
The fourth scenario is the same policy as the third scenario, but the off-street parking prices are adjusted instead of on-street parking prices. The demand-responsive pricing for garage parking sets the on-street parking prices identical to the benchmark model and the parking garage fee is adjusted, so there will be always a free spot in every parking garage.
Finally a social-optimum scenario is defined. In this case parking fees are not considered, because it is assumed that every agent has enough money to pay any fee. The goal of this scenario is to minimize the time cost in the system.
Figure 4.12 shows the time needed to search a parking space a the different destinations.

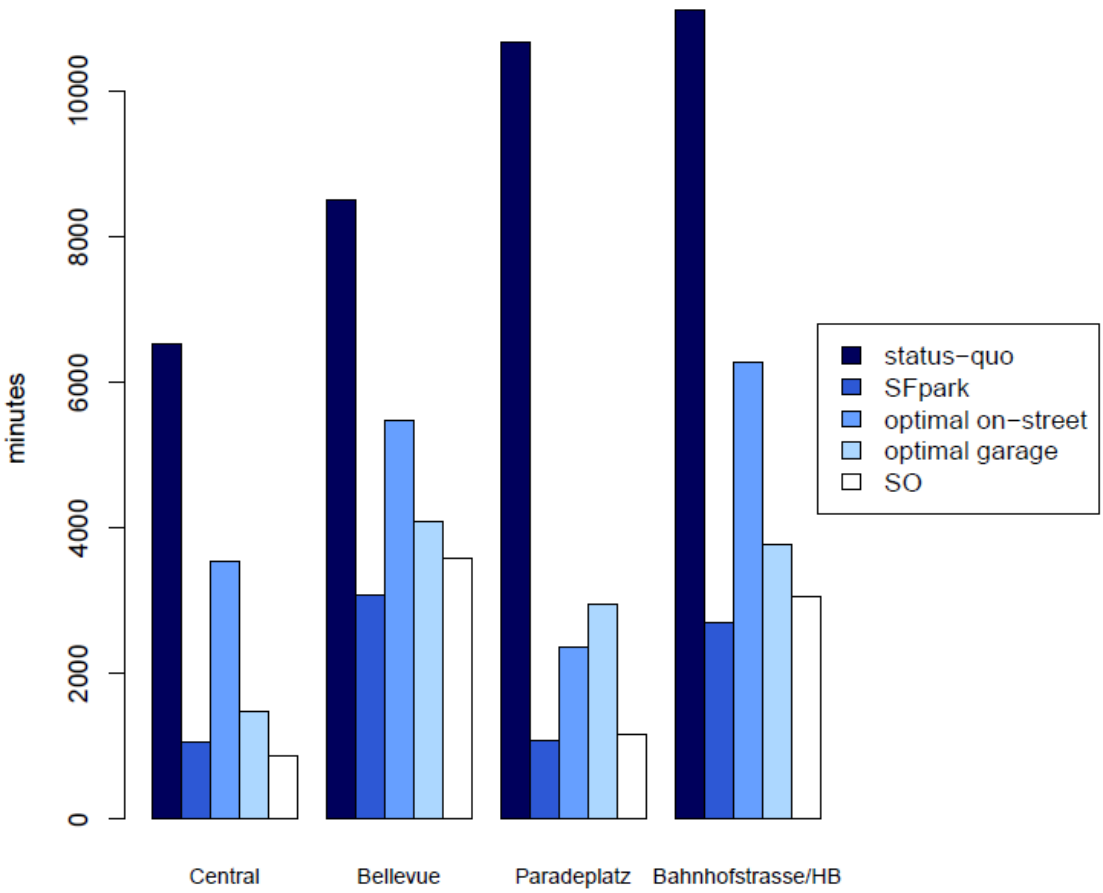


Figure 4.12: CGE-Model Zurich: Total Search Time by Destiniation [2]

As can be seen, the current parking model in Zurich is always worse than the alternative scenarios, with significantly more time spent searching for parking spaces. It is interesting that one needs more time to search a parking in Bahnhofstrasse, although there is more traffic in Bellevue. That is because the valuation by household from the inflow nodes near Bellevue is higher than by Bahnhofstrasse. In other words, the drivers who search for a parking space near Bellevue tend to come from Utoquai, because it is near, and they can afford a higher price. Another observation is that the optimal on-street parking policy is always worse than the optimal garage policy. This can be explained by the fact, that the time spent in the garage searching for a free spot is not considered in Figure 4.12, because the cars driving in a garage are not congesting the traffic on the street.
In conclusion, the model shows that it is possible to reduce the overall flow in the network by applying a demand-responsive parking policy like SFpark. The question arises, why the city of Zurich and other cities are not investing more in finding alternatives to the status quo. To implement new parking policies, such as the ones considered in this report, is very demanding. Since most of the on- and off-street parking spaces are privately owned, and therefore it is impossible to regulate fees through a citywide parking policy. However, it might be possible to encourage parking owners to apply adjusted prices, especially in parking locations which further away from the city center. This would attract some drivers to park their car away from the center, using parking spaces with a low occupancy rate and thus decreasing the traffic congestion in the rest of the city.

4.5 Conclusion

Missing parking availability information has severe consequences. It does not only affect the drivers who cruise for a free parking spot, but also negatively influence the environment

and other traffic participants. Drivers lose time which they might use more efficiently, and they waste fuel. The increased energy consumption pollutes the environment, and the cruising for parking leads to further noise pollution. Furthermore, traffic is increased, and thereby also the risk of traffic congestions and traffic accidents are higher.

The negative impacts of missing parking availability information and the consequential cruising for parking is versatile, but there are multiple approaches to address the issue. The introduced solutions in this report all consider the problem from different perspectives: ParkU provides additional parking spaces by enabling parking space owners to monetize their parking spaces. Parking space availability can be checked via app or website and parking spaces can be reserved in advance. parkITsmart uses a information crowdsourcing approach, but also considers the perspective of parking inspectors. The park & ride guidance system in Donostia-San Sebastian reduces traffic in the city by providing and guiding drivers to the best parking option before entering the city and builds upon existing infrastructure. SFpark, as the most expensive solution to implement introduced here, requires a lot of effort, infrastructural changes and governmental spending, but allows for dynamic price adjustments. These dynamic price adjustments allow for optimizing the occupancy rate of parking spaces in San Francisco. Finally, the CGE-Model for Parking Spaces in Zurich shows that the current situation in Zurich is not ideal and can be improved. It also shows that evaluating the parking situation in a city is highly dependent on city specific variables, and that the situation in each urban area must be considered individually.

While it is clear that the situation in many urban areas should be addressed, one cannot apply the findings of one city to another. Cities must carefully evaluate the available options and the consequences of implementing them. In order to evaluate different options more efficiently, further work is required in creating frameworks allowing individual cities to compare valid solutions for their specific case.

Bibliography

[1] Iñaki Baro: *Park & Ride Guidance System*, Civitas, Archimedes Deliverable No. 75, Donostia-San Sebastian, 2011.

[2] Anne - Kathrin Bodenbender: *A CGE - Model of Parking in Zurich: Implementation and Policy Tests*, M.A. thesis, ETH Zurich, Switzerland, 2013.

[3] J. Cukas: *Deutsche helfen Zürchern, einen Parkplatz zu finden*,Tages Anzeiger, `http://www.tagesanzeiger.ch/zuerich/stadt/Deutsche-helfen-Zuerchern-einen-Parkplatz-zu-finden/story/19125477`, February 2, 2013 [October 10, 2015].

[4] Anthony Downs: *Still Stuck in Traffic: Coping with peak-hour Traffic Congestions*, Washington, DC: The Brookings Institution, 2004.

[5] Livio Hobi: *The Impact of Real-time Information Sources on Crowdsourced Parking Availability Prediction*, M.A. thesis, University of Zurich, Zurich, 2015.

[6] Fabian Hofstetter: *A Public Parking Management System in Zurich*, M.A. thesis, University of Zurich, Zurich, 2015.

[7] Eren Inci, Jos v. Ommeren & Marijn Kobus: *The External Cruising Costs of Parking*, Tinbergen Institute Discussion Paper, Amsterdam, 2015.

[8] justpark.com, `www.justpark.com`, [October 10, 2015].

[9] Samuel Liniger: *Parking Prediction Techniques in an IoT Environment*, M.A. thesis, University of Zurich, Zurich. 2015.

[10] Jos N. van Ommeren, Derk Wantink & Piet Rietveld: *Empirical Evidence on Cruising for Parking*. Transportation Research, vol. 46(1), pp. 123-130, 2010.

[11] *Parking Panda*, `www.parkingpanda.com`, [November 12, 1015].

[12] *Parku: the parking App*, `www.parku.ch`, [November 12, 1015].

[13] Gregory Pierce & Donald C. Shoup: *Getting the Prices Right: An Evaluation of Pricing Parking by Demand in San Francisco*, Journal of the American Planning Association, vol. 79(1), pp. 67-81, 2013.

[14] Donald C. Shoup: *The High Cost of Free Parking*, Updated Version. Chicago: Planners Press, 2011.

[15] Donald C. Shoup: *Cruising for Parking*, Transport Policy, vol. 13(6), pp. 479-486, 2006.

[16] Donald C. Shoup: *Cruising for Parking*, Access, vol. 30, pp. 16-22, 2007.

[17] Hartmut H. Topp: *Parking Policies to Reduce Car Traffic in German Cities*, Transport Reviews, 13(1), pp. 83-95, 1993.

[18] Christos Tsiaras, Livio Hobi, Fabian Hofstetter, Samuel Liniger and Burkhard Stiller: *parkITsmart: Minimization of Cruising for Parking*, In Proc. of the 24th International Conference on Computer Communication and Networks (ICCCN), pp. 1-8, 2015.

[19] William Vickrey: *The Economizing of Curb Parking Space*, Traffic Engineering, vol. 29(1), pp. 62-67. Chicago, 1954.

Chapter 5

Watermarking in Sensor Data Sets

Melina Mast, Romana Pernischova, Manuel Schlegel

Watermarking is no new concept. It has been used for decades as identifying images or patterns in postage stamps and passports, as a signet on letters and many more applications. The digitalized form of watermarking is not as widely known. However it can be used for several applications like copy control or transaction tracking. Another way of using these digital watermarks is the embedding of a watermark into sensor data sets that have been produced by Wireless Sensor Networks or short WSN. Nonetheless, because of the extreme limitation of WSN resources, such as computational power, battery life, and memory capacity, security solutions need to maintain the trade off of security and cost. Embedding and extracting a watermark from a packet requires often less computation than encrypting the whole packet. For this reason, three different methods for authentication are presented and briefly discussed in this report: Authentication Methods for Self-Nodes, Self-Identifying Sensor Nodes, and Copyright protection using Watermarks.

Contents

5.1 Introduction **99**
5.2 Digital Watermarking **99**
5.2.1 Introduction to Digital Watermarking 99
5.2.2 Evolution of Digital Watermarking 100
5.2.3 Applications of Digital Watermarking 100
5.2.4 Advantages and Disadvantages of Digital Watermarking 101
5.3 Wireless Sensor Networks **102**
5.3.1 Introduction to Wireless Sensor Networks 102
5.3.2 Wireless Sensor Networks Applications 103
5.3.3 Security and its Issues in Wireless Sensor Networks 103
5.4 Methods for Authentication **105**
5.4.1 An Authentication Method for Self Nodes Based on Watermarking in Wireless Sensor Networks 105
5.4.2 Self-Identifying Sensor Data 110
5.4.3 Copyright Protection in Wireless Sensor Networks by Watermarking 114
5.4.4 Comparison of Methods for Authentication 118
5.5 Summary and Conclusion **118**

5.1 Introduction

We are daily confronted with objects that contain a watermark, such as bank notes, post stamps, your passport and even your DVD [1].
Watermarks have been used for decades on currency and other government documents to discourage counterfeiting. For example, on the paper of a bank note you can see an identifying image or pattern that appears as various shades of lightness or darkness when viewed by transmitted light or when viewed by reflected light, atop a dark background, caused by thickness or density variations in the bank note. In the 1990's with the advances in computer technology, digital watermarking techniques to protect copyrighted content became more important, as the possibility for unlimited copying of digital data and their distribution over the newly developed World Wide Web became reality. Therefore, watermarking methods were developed for different applications that are also presented in this paper like digital copyright protection [1] [3] [15].
Watermarks also find their application in Wireless Sensor Networks (WSN). None of the well-known digital watermarking methods are directly applicable to WSNs. Regardless, because the sensor data is unstructured, the usage patterns for communication are unique and traditional watermarking techniques are too complex for the limited computational capability and storage capacity in WSNs [9].

However, there exist solutions for watermarking in WSNs and we will present you a selected set of methods. Firstly, in Section 5.2, we give you an introduction to Digital Watermarking, its evolution, applications and challenges. Section 5.3 presents WSN, their applications and discusses security issues. Section 5.4 provides three different methods for authentication purposes: Authentication Methods for Self-Nodes in Chapter 5.4.1 shows an example how wireless sensor nodes can be authenticated with a watermark [3], Self-Identifying Sensor Nodes in Chapter 5.4.2 deals with the provenance problem of wireless sensor data [9] and Copyright protection using Watermarks in Chapter 5.4.3 proposes a watermark algorithm for protecting the copyright of sensor data [10]. In the end of the report, that is in Section 5.5, we will summarize and conclude the findings.

5.2 Digital Watermarking

Digital watermarking is a possibility to embed information to digital data without being noticed. An aim of digital watermarking is to ensure that the copyright of the author is protected. Further, a main purpose of watermarking in WSNs is to transmit information via insecure channels [11] [3].

In this chapter we give a brief introduction to digital watermarking. Further, we describe its evolution and possible applications. At the end, we discuss the advantages and disadvantages that digital watermarking implies.

5.2.1 Introduction to Digital Watermarking

Since the 1990's the interest in digital watermarking has increased due to the advances in computer technology and the development of the Web. These technologies offered a possibility to copy and to distribute copyrighted material to nearly everywhere in the world. Therefore, cryptography possibilities have been developed to protect the distribution of copyrighted content. Once the content has been decrypted, all protection is lost. Hence, watermarking is often used to compliment cryptography to provide protection after decryption.

However, a question is if watermarks are interesting from a business perspective. From a security perspective, many watermark security solutions are considered as ”weak”. Nonetheless, ”weak” security is often preferable to no security [3] [11] [8].

5.2.2 Evolution of Digital Watermarking

Digital watermarks have been deployed for several decades. In 1954, the first known patent for electronic watermarking named ”Identification of sound and like signals” was granted [8]. In the patent a method for embedding an identification code into music to prove ownership is described. This was the beginning of electronic watermarking.
Owner identification has been pioneered by the Muzak Corporation [8]. The system they used encoded identification information using Morse Code. The system was used until the early 1980’s, when their interest in identifying the music they owned decreased. After that, each DiVX DVD player contained watermarks that supported transaction tracking to prevent piracy.
Some companies tried to embed signaling for device control. Further, a company named Musicast Inc. proposed to distribute music to businesses [8]. Their patent describes embedding a low frequency 30Hz signal at transmission point which allows the receivers to remove advertisements. Another company has received a patent in 1976 for a video watermark for television applications. In 1989 a patent for a ”Method and apparatus for in-band, video broadcasting of commands to interactive devices” was issued[8].
These companies offered watermarking solutions for a number of different applications. This includes interactive television, interactive toys and advertisement monitoring.
In addition to the watermarking activities until the 1990’s, several new applications have been developed, because of the decreasing interest to prevent the distribution of copyrighted data. Many companies have used watermarking for several years now, and they are considered as evidence that watermarking is commercially viable. Some of the applications are described in more details in the upcoming section [8].

5.2.3 Applications of Digital Watermarking

The applications of watermarking can be classified as copyright control like owner identification, proof of ownership, transaction tracking, and copy control.
The potential applications are described in the following chapter [8].

5.2.3.1 Transaction Tracking

Transaction tracking, also known as fingerprinting, is a unique watermark that is embedded into each copy of a work. Normally, the watermark identifies the legal recipient of the copy. Further, it can be used to trace the source of illegally distributed content. An example for transaction tracking is the watermark that is implemented by DiVX. However, it is known to be vulnerable to collusion attacks. In a collusion attack an adversary has several copies of the same work and each has a different watermark. The adversary uses the copies to obtain an approximation of the watermark-free work, the original. Most existing watermarks can be removed using fewer than 20 copies. Therefore, an adversary could produce watermark-free copies. Nevertheless, the system might still be worth it because the adversary needs the knowledge to perform these attacks, which can already be seen as some prevention of piracy attacks [8].

5.2.3.2 Proof of Ownership

The original interests of the Muzak company was to be able to differentiate between their own and similar recordings. Another ambition is to prove ownership in a court of law.

A problem is that it is possible for adversaries to make it appear that all distributed copies of a work contain a watermark embedded from the adversary. However, a solution for this problem would be a cryptographic link between the watermark and the original work. Therefore, it is technically possible to prove ownership using watermarks [8].

5.2.3.3 Copy Control

There are two challenges in implementing a copy-control system.
The technical problem is that everyone must be able to detect the watermarks. So far, the current technology only provides weak security.
The political problem is that equipment manufacturers must include watermark detectors in their recording devices. For the equipment manufacturers, watermark detectors just add cost and no value. Actually the equipment even loses value, because some consumers would like to make illegal recordings. A solution could be to force manufacturers to include detectors by law [8].

5.2.3.4 Authentication

A digital signature can be embedded as watermark for authentication in a work. This is an advantage for legacy systems that is explained in Section 5.2.3.5. There has been a concern that embedding a signature alters the work. However, we will address this concern in Section 5.4 [8].

5.2.3.5 Legacy System Enhancement and Database Linking

Watermarking pays a valuable role in improving the functionality of legacy systems. Especially while maintaining compatibility with deployed devices. For example, Philips has demonstrated an audio watermarking technology for music, which can be used to identify a song. They are using the microphone that is present in many PDA's (personal digital assistant) to decode and identify a song. If the PDA has a wireless Web connection, it can link the song to associated sites, which may have additional information to the song. Robustness, fidelity and payload requirements are important in legacy system enhancement and database linking [8].

5.2.4 Advantages and Disadvantages of Digital Watermarking

In the following, we discuss whether watermarking really offers practical solutions and if it is a worthwhile topic of research. Besides, we illustrate the advantages and disadvantages of digital watermarking.
Watermarking must maintain the accuracy of the underlying work besides transmitting a message. The fidelity and the robustness constraint often conflict. Robustness defines how well a watermark stands up to transformations of data [9].
We consider the watermark communications channel to exist in an adverse environment. An example is the usage for copyright purposes. There is often a strong intention to remove the watermark. Some examples for attacks are collusion attacks, ambiguity attacks, copy attacks, gradient descent attacks, and sensitivity. In the last years there has been a lot of progress in the development of attacks.
Much of the effort in developing watermarks has focused on fidelity, robustness and economic constraints. Many design choices are available but no one signified a breakthrough. If you want to use watermarks for large-scale deployment, *e.g.* a DVD copy control application, the watermark detector has to be very inexpensive to be usable. Furthermore, it is important that the code is computationally efficient and robust. Besides, the conflict between the fidelity and robustness constraints when embedding a watermark has to be

optimized.
There are watermark applications that do not need to be secure (*e.g.*, watermarks that link media to the Web).
Further, even weak security is valuable in many business environments. Nevertheless, it is still not known whether a watermark system can be designed so that it permits public detection of the watermark and that it prevents an adversary from removing the watermark. Sensitivity analysis and gradient descent attacks are threatening any watermarking systems when the detector is publicly available.
In conclusion, even if watermarking has some unresolved issues, it will still be used by businesses, because weak security is better than no security at all [8].

5.2.4.1 Application of Digital Watermarking in Wireless Sensor Networks

As we have seen, there is a vast array of digital watermarking techniques. However, none of them are applicable in environments where WSNs are used. The reason for this is that sensor data is, unlike watermarks used in databases, unstructured. Besides, the usage patterns are unique for every user community, and because watermarks should be imperceptible in usage and robustness, they must be adopted to this environment [9]. Before describing methods of watermark usage we will talk about WSNs in general in the next chapter.

5.3 Wireless Sensor Networks

A WSN is built out of interconnected sensor nodes, which communicate using radio frequency, operating on very limited hardware to deliver data about the environment that they are deployed to. The hardware limitations cause great challenges not only in computational capabilities, but also in security options. Encryption with symmetric keys takes a severe amount of computational power, which would then lead to a fast dying of nodes and make the network useless too fast to deliver the desired outcomes [2] [4] [5].

In this chapter, we would like to give you a brief introduction to WSNs, its limitations and challenges. Talking about a few different applications should give you a taste, where and when WSNs are deployed. The real focus of this chapter however is on security and its issues of WSNs, which bring different challenges.

5.3.1 Introduction to Wireless Sensor Networks

A WSN is a network consisting of sensor nodes, which communicate wirelessly with each other and with a base node. The nodes are small, autonomic and programmable and different sensors can be attached.
This leads us to the topic of node limitations. Typically, a sensor node has a four to eight MHz processor, four kBytes of RAM, 128 kBytes of flash memory and a radio frequency of 916 MHz [7]. The life of such a node is dependent on the battery that it is operated by. For this reason, algorithms, transmitting and any other activities must be very battery efficient, so that the life of a node can be extended as far as possible. Of course this brings many different and important challenges with it.
Another limitation is in the network itself. This limitation is comparable to limitations of a mobile ad-hoc network with additional lack of infrastructure and also being forced to rely on not secured wireless connections.
Physical limitations like exposure to vandalism, are present due to the types of applications that WSNs are used for. Normally, WSNs are deployed in public or environmentally

dangerous places. One way to improve these limitations, is to use tamper proof materials for sensor nodes, which would then lead to higher cost. Depending on the application a higher cost cannot be taken into consideration. For example, when deploying a WSN into a forest fire, one does not expect to collect the unharmed sensor nodes once the fire has been extinguished. The nodes are deployed only for the safety of the firemen and that is why the sensor nodes need to be as cheap as absolutely possible. In another situation, it would be a different matter, for example in a health care application, where the data must be trusted at all times.
In conclusion, sensor nodes are restricted in computation capability, power supply, storage capacity and radio transmission range because of being extremely constrained in hardware [2] [7] [5].

5.3.2 Wireless Sensor Networks Applications

There are many applications for WSNs. We have chosen two different examples to present to the reader very briefly.
[12] presents a project, where vulcanologists and WSN researchers collaborate to use a Wireless Sensor Network to monitor an active volcano. The goal of deploying a WSN in an active volcano was to use sensor nodes to collect seismic and infrasonic signals. Infrasonic refers to low-frequency acoustic signals. The group was dealing with radio links being lossy and frequently asymmetrical and also that the low-cost crystal oscillators on each node have low tolerances, which lead to varying clock rates across the network. A single missed or corrupted sample could then invalidate an entire record of data. Differences, in for example the sampling rates, would then also frustrate analysis.
For this project the group developed a reliable data-collection protocol, which they called Fetch, to be able to retrieve the data from the nodes using a multi-hop network.
Deploying a WSN in different health care applications raises many security questions which lead to stringent security and privacy requirements. In [13] the authors describe these requirements and already existing solutions along with some specific applications.
The goal of deploying a WSN in any health care application is to enhance the service, which would include emergency response, in-hospital communications, out-hospital monitoring or also environmental monitoring.
Security questions and challenges do not arise only in health care applications, but in any applications where WSNs are involved. The next chapter introduces these challenges and also the issues and what attempts and directions have been looked at to solve these.

5.3.3 Security and its Issues in Wireless Sensor Networks

Depending on the application it is important that the WSN is secure. This means that the data that is being transmitted, the nodes, and also the base station can be trusted. For this case we can simply state some security goals, as in [5] and [14].
Confidentiality stands for a requirement that ensures that sensitive information is protected and is not revealed to other parties. Such a requirements is necessary in WSNs so that the information that is traveling between nodes or a node and the base station is protected from disclosure. An attacker that has appropriate resources could eavesdrop on the conversation between these nodes or a node and the base station. It would be possible to overhear critical and sensitive information and the attacker could then use such information for illegal purposes, such as blackmail, or then attempt to infiltrate the network even further to be able to overhear the entire conversation. Could an attacker through eavesdropping steal encryption keys it could have disastrous consequences.
Integrity is another goal or requirement for WSNs. Integrity addresses the issue of a message being modified when exchanged through an unsecured link. In most applications the

usage of inaccurate or false information could ruin the whole work or have catastrophic consequences. For this reason integrity controls have to be implemented to ensure that information was not altered.
Authentication is introduced to be able to identify participants in a network so that one can know about legitimate users and intruders. In a WSN such security requirements must be ensured at any time. The network has to be able to tell if all its components are legitimate and that no malicious nodes have entered the network.
Availability is the last requirement which we would like to address. Availability deals with the issue of services and information being accessible and available at any given time. For real time applications, like monitoring a patient at a health care facility, availability is crucial and also critical. The system has to be able to fill the gap in the case of a node unexpectedly leaving the network. This node's duties have to be assigned to some other node so that this requirement is fulfilled. It also applies to nodes that have special roles, like aggregation of data or even the base station [5] [14].
The following list from [4] provides requirements for WSNs on secure networks. Such requirements need to be integrated into the system completely and not viewed as standalone features, because a separation would not yield the best possible results.

- Key establishment and trust setup
- Secrecy and authentication
- Privacy
- Robustness to communication denial of service
- Secure routing
- Resilience to node capture
- Network Security Services such as secure group management, intrusion detection and secure data aggregation

Established security solutions, mainly cryptography, stenography, and Physical Layer Secure Access have been considered as a solution to security problems in WSNs. However, they are confronted with different challenges once these are to be applied to an ad-hoc-like wireless network like a WSN. These issues are also further discussed in [6].
Cryptography is a well-established technique, which is designed for wired networks and is not feasible for an application in Wireless Sensor Networks. When using encryption of the data that is to be transmitted, each node would have to transmit additional bits which would require more processing, memory and battery power. One would also have to deal with delays, jitter and greater packet loss. Due to these problems cryptography is not well applicable for WSNs. When seriously considering encryption one would also have to solve the issue of key distribution and management. An adoption of pre-loaded keys is not an option when wanting to add new nodes to a network.
Stenography strives to hide the existence of the message, instead of the content, like cryptography. The goal is to embed the message into multimedia data and make the message look ordinary. WSNs and its data sets are not really related to stenography and due to the extreme limitations of resources it is not really relevant here.
Physical Layer Secure Access could be provided, for example by frequency hopping, which can be achieved through a combination of parameters like hopping set, dwell time and hopping pattern. The expense of memory, processing and energy resource would not be as great as with cryptography [6].
Just like in networks in general there are several types of attacks in WSNs that can be divided into two groups of active and passive attacks. The most common attack is a

Denial of Service attack. There are many different forms of this attack depending on the attack point and also on the purpose behind the attack. In [5], [6] and [13] we find descriptions of different forms of attacks on WSNs. Since we will mainly be concerned with methods for authentication in WSNs, we will not discuss this matter any further.

5.4 Methods for Authentication

In this section we present three examples for authentication methods using watermarking. Each of the methods is summed up, described and is compared in performance and security. Further, we discuss the three methods in terms of field of use, advantages and disadvantages, and their boundaries.

5.4.1 An Authentication Method for Self Nodes Based on Watermarking in Wireless Sensor Networks

A WSN can be widely used in various fields. Since a lot of them are deployed outside, where they are not monitored by humans, an authentication method becomes important to ensure the data security.
For example, an enemy may deploy some nodes to present to the nodes within the network, which are meant by using *self nodes*. The willfully deployed nodes may send some fake data to the self nodes or try to perform attacks on the network. The fake data may lead to error decisions if it is not recognized but instead transmitted and aggregated with the correct data.
In order to prevent such error decisions, authentication of the self nodes should be used. An authentication method based on public key cryptography is not usable, because of the limited computing capability, storage, bandwidth and power of WSNs. Hence another method should be used. Compared to the public key cryptography, the watermarking based authentication has the advantage that it is less complex in terms of computation. Besides, the watermarking technique is invisible to enemies.
In some WSNs the security of the whole network is defected even if just one node is not reliable. The described method was developed in consideration of this problem. It is a watermarking based authentication technique that should ensure the trustworthiness of data transmission by authenticating the sending nodes identity.
Before the data is transmitted, the identity of the node is integrated into the data and stored until the buffer is filled. The node that receives the data can find the embedded watermark with particular keys and parameters. Only if the data is correctly watermarked it can be trusted. Further, only if the data is reliable it can be stored and transmitted. The data that is not identified as reliable should be discarded. Thus, the malicious data can be blocked in the network.
By this model, the nodes can be divided into two classes, because of their tasks: The data collection/embedding nodes and the data detection/transmission nodes. The embedding and detection algorithms cannot be very complex, because of the limited computational capability of WSNs. The watermarking model is visualized in Figure 5.1. The figure shows the embedding and detection process. First the data is embedded. Next, the watermark is embedded till the provided buffer is filled. If this is the case, the data can be send to the receiving node. The receiving node detects the watermark and afterwards decides if the data should be stored or transmitted or if it should be discarded [3].

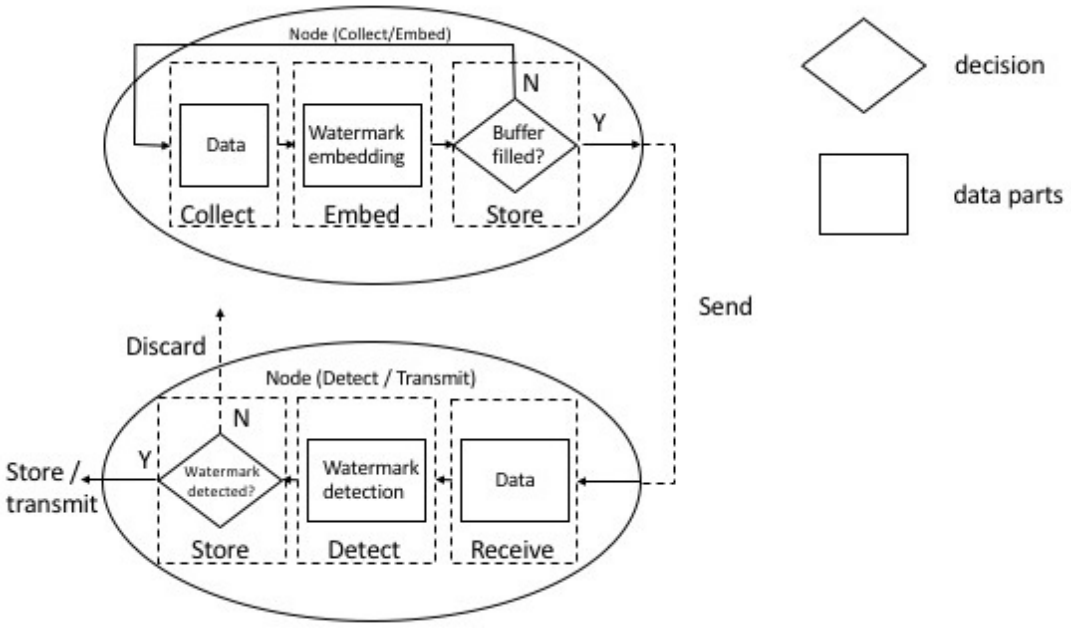


Figure 5.1: Authentification model [3].

5.4.1.1 Watermark Generation

The watermark is designed to identify the sending node and the sending data. In this method a key and the data collection time is used to generate the watermark as illustrated in Listing 5.2.
The data packet can be represented as: *Packet = (head, send-data). Head* is the packet head which includes information about route and packet length. *Send-data* is the data content which should be sent. The elements of the collected data are defined as: $s_i = (t, d_1, d_2, ..., d_n)$ where i is the work cycle of the node and t is the collecting time of the data. Further, $d_j (j = 1, ..., n)$, where j denotes the items of the collected data. A packet can consist of various head and data within k cycles $(s_1, s_2, ..., s_n)$.
In each data element s_j, a key K and the collection time of the key is s_j, t inserted into a hash function. $h_i = Hash(K, s_i, t)$ is the hash function that is calculated. One bit of the watermark $WM[i]$ is calculated by the exclusive or of the w bits of the $MSB[h_i]$. The final watermark WM contains the collection of all $WM[i]$.
First, the hash function is calculated and then the Xor of the w leftmost bits is calculated. The result for $WM[i]$ is either a 1 or a 0. The results are embedded as parts of the watermark. If the detecting algorithm is successful, it calculates the same results for $WM[i]$. These results are compared to each other and decided if the data is accepted. If the bits are equal the data is accepted. The tolerance level for the number of aloud wrong bits can be decided [3].

5.4.1.2 Watermark Embedding Algorithm

The change to the data, because of the data embedding should be minimized, the embedding is designed at the least significant bit of the data. Therefore, the rightmost bits. A random value of each data item s_i, d_i was calculated by inserting the collecting time, its most significant bits and the key are the input for a random function. A parameter v was selected, which defines the proportion of watermarked data. Afterwards, the watermark is embedded into the data items that have a random value dividable exactly by v. Further,

Watermark Generation Algorithm
Input: sending-data s_i, key K, number of bits to select w;
Output: watermark;
Steps:

Generate WM(s_i, K, w)
1) $h_i = Hash(K, s_i, t)$
2) $WM[i] = Xor(MSB(h_i, w)$
//select w bits from $MSB(h_i, w)$ *and calculate the exclusive or of the w bits.*
3) return $WM[i]$

Listing 5.2: Watermark generation [3].

Watermark Embedding Algorithm
Input: send-data, key K, number of bits to select w, parameter v, parameter c;
Steps:

Embed WM(send-data, K, w, v, c)
//K, w, v, and c are shared by the self nodes
1)*Generate WM*(s_i, K, w)
2) for each s_i, d_j *do*
3) g = *Rand(K,* s_i*, t, MSB(*s_i*,* d_j*)) //calculate the random value for the data item*
4) if (g mod v == 0) then
5) x = *g mod c //calculate the embedding location*
6) set the xth *LSB(*s_i*,* d_j*) do* $WM[i]$
7) end for

Listing 5.3: Watermark embedding [3].

the random value is used to calculate the embedding location, the xth bit. The xth bit will then be set to $WM[i]$. The embedding algorithm is shown in Listing 5.3 [3].

5.4.1.3 Watermark Detection Algorithm

The data receiving nodes have to extract and verify the watermark. The received data is represented as *Packet = (head, received-data).* It contains $(r_1, r_2, ..., r_n)$ where the content is the same as in s_i.
The watermark is detected successfully if the detection rate is bigger than the threshold level t as you can see in line 12 in Listing 5.4. Else the data will be considered not watermarked and the algorithm will return false. The threshold level can be set considering several aspects like the potential attacks, the robustness of the algorithm or the accuracy balancing. The self nodes share three parameters: the data, the key parameter, and w which is used to generate the watermark again that will be compared to the watermark embedded in the received data. The threshold is to be set according to the security requirements. The detecting algorithm is shown in Listing 5.4 [3].

5.4.1.4 Performance Analysis

The Key Laboratory of Medical Image Computing (Northeastern University) and the Ministry of Education from the Northeastern University designed experiments for testing the algorithm's robustness and security. They made tests with two values for v, 3 and

Watermark Detecting Algorithm
Input: received-data, key K, number of bits to select w, parameter v, parameter c;
Steps:

DetectWM(send-data, K, w, v, c)
// K, w, v and c are the same as in the embedding Algorithm in Listing 5.3
1)$totalcount = matchcount = 0$
2)$Generate\,WM(s_i, K, w)$
3) for each r_i, d_j *do*
4) $g = Rand(K,\ r_i,\ t,\ MSB(r_i,\ d_j))$
5) if (g mod v == 0) then
6) x = g mod c //calculate the embedding location
7) totalcount = totalcount + 1
8) if(the xth $LSB(r_i,\ d_j)$ *is* $WM[i]$ *then*
9) matchcount = matchcount +1
10) end for
11) detectrate = totalcount / matchcount
12) if(detectrate > t) then
13) return true
14) else
15) return false

Listing 5.4: Watermark detecting [3].

5, and for testing the robustness if the data was summarized or sampled. The sensor data was collected by the Intel Berkeley Institute that was adopted in the experiments, including data about temperature, humidity and illumination [3].
The security of the algorithm concerns the ability of adversary nodes to successfully send packets to self nodes without being detected as malicious nodes. An adversary node tries to be seen as a self node and forges some data packets to send it to other nodes. However, as the adversary node cannot know the secret key K and the other shared parameters, he cannot forge a correctly watermarked data packet. Hence, the detection rate for the forged data packets should be smaller than the threshold. Therefore, the forged data can be found by comparing the detection rate and the threshold. The detection rate was always around 100% for watermarked data packets and around 50% for not watermarked packets. In this case, the difference in the detection rate is not significant for different values of v. The results are illustrated in Figure 5.5. The figure shows the result of the experiment once using 3 and once using 5 for the parameter v. The results indicate that the higher the parameter v is the better is the detection rate. Besides, there must be a level from where v achieves a detection rate of 100% and this value must be between 3 and 5 [3].

Data are often fused during transmission to decrease the transmission data. Therefore, the watermark needs to be ensured not to be removed or damaged. To test the robustness the effect of sampling and summarization on the data were tested. The effect is illustrated in Figure 5.6. We can see that the sampling does not affect the detection rate. However, summarization leads to a decreasing detection rate when the proportion of summarized data is increasing. Further, the detection rate is higher with a large value of the parameter v [3].

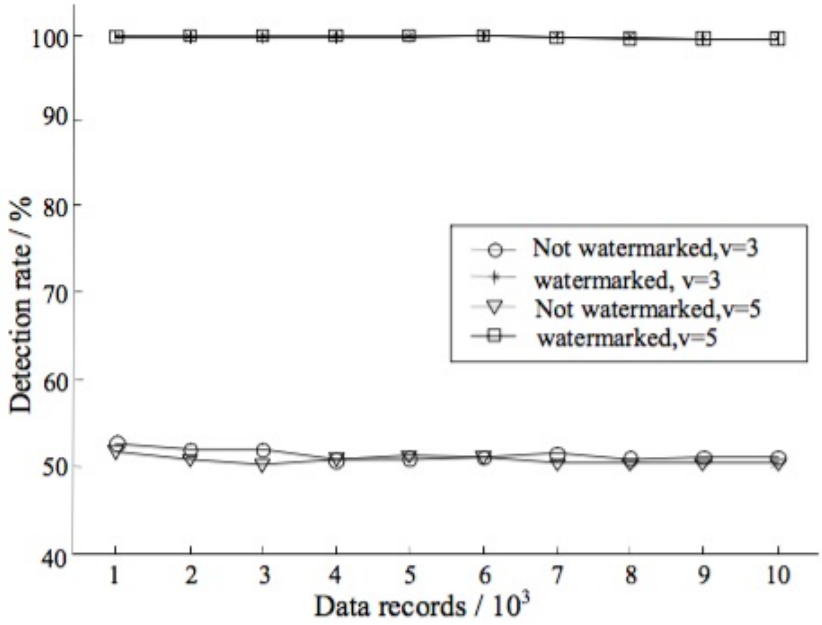


Figure 5.5: Detection Rate Comparison of Digital Watermarking [3].

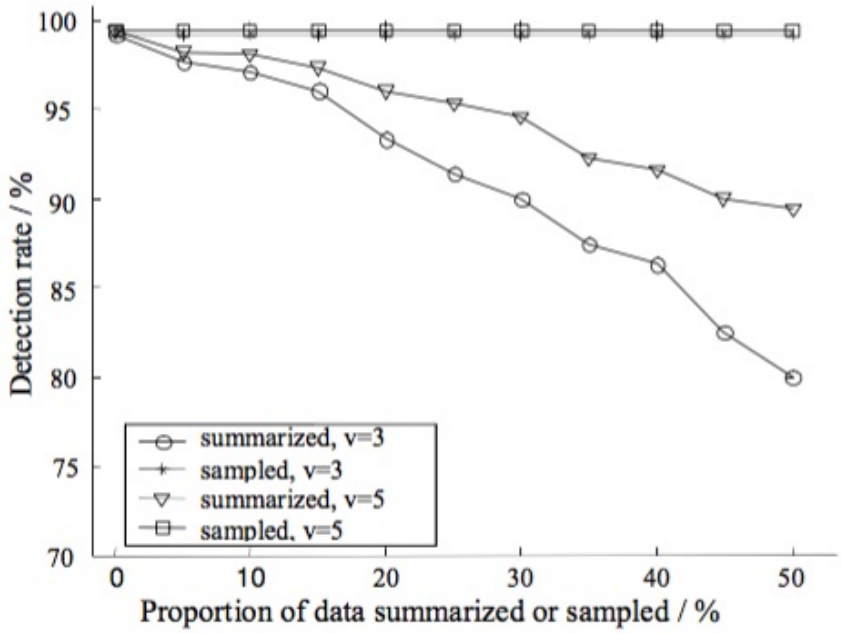


Figure 5.6: Proportion of data summarized or sampled in % [3].

5.4.1.5 Findings of First Authentication Method

WSNs can potentially be used in various fields. However, because of the limited resources, they are confronted with some challenges on network security. By using the algorithms, the data from malicious nodes can be refused and be blocked. Further, the algorithm allows you to adapt the security to your environment. Yet, the limited resources, aspects like energy consumption, algorithm complexity and network performance should always be concerned [3].

5.4.2 Self-Identifying Sensor Data

Sensor data sets for public-use are a useful scientific resource. Conditions of use of these data sets usually include that users cite their contributions. Unfortunately, data is used and passed around a lot, which may lead to a loss of this information. The source of the data sets, the *provenance*, is easily disconnected from their content, even without malicious intent. To address this problem we will introduce a way for sensor data to directly associate the source information with the data. The solution is *semi-fragile*, which means that it is capable of withstanding many, but not all transformations [9].

5.4.2.1 Informal Problem Description of Self-Identifying Sensor Data

The technique we will describe in the following, embeds provenance information in the lower-order bits of data points in a data set. For *one-bit checking* the question the technique wants to answer is: "Is a provenance mark m embedded in the given data set DS?". Two of the lower-order bits of every data point are used as *check bits*. One of the check bits helps to determine the embedding parameter. The other one is a hash of the encoded provenance mark concatenated with the more significant bits of the data point. For *blind checking* the question is: "Which provenance mark was embedded in the given data set?" To check this, the provenance mark is separated into smaller pieces, so called *provenance pieces*, which replace the least significant bits of each data point one of these provenance pieces. A bit of the provenance mark may appear in more than one provenance piece to increase robustness to truncation and rounding. The significant bits determine which provenance piece replaces the least significant bits. During the retrieval process, the least significant bits are examined to retrieve the "best guess" at the provenance mark. This guess can be tested using one-bit checking. If the one-bit check finds out that it is incorrect, the retrieval process allows limited search of possible provenances marks, where more likely marks are considered first [9].

5.4.2.2 Embedding a Provenance Mark in a Data Set

The embedding of a provenance mark in a data set goes as follows. We assume that each data set contains insignificant bits. These insignificant bits are located on the right side (the last bits) of the data, while the significant bits are on the left side. These bits can safely be manipulated within error bounds. The algorithm now embeds a provenance mark into the data set by encoding metadata into the insignificant bits of each data set. A data set that has received an implementation of such a provenance mark is called an *uncorrupted annotated data set*, which contains data points whose insignificant bits have been replaced with meta data and are now referred to as *uncorrupted annotated data points*. Figure 5.7 represents the structure of an annotated data point [9].
The provenance pieces are generated using the formula in Equation 5.8 [9].
Where L_m is the assumed fixed length of the data set, N_{pp} is the number of provenance pieces, s is the amount of significant bits, $k = hash(N_{pp}, s)$ and i is the i-th bit of the k-th provenance piece, denoted pp_i^k [9].

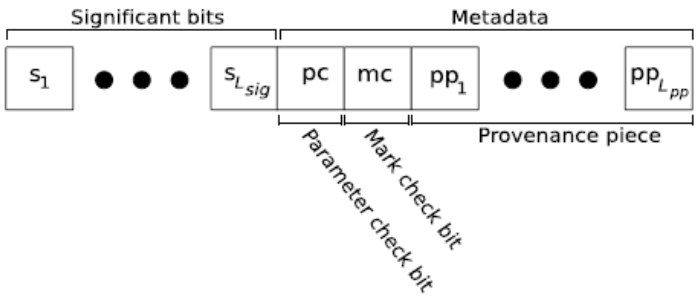


Figure 5.7: Structure of an annotated data point [9].

$$\mathrm{pp}_i^k = \mathrm{m}_j \text{ where } j = \left(k\frac{L_m}{N_{pp}} + i\right) \mod L_m.$$

Equation 5.8: Provenance piece generation [9].

An example result of the formula is shown in Figure 5.9 [9].
In the figure, L_{pp} is the length of each provenance piece. More importantly, the insignificant bits have been substituted with a parameter check bit, a mark check bit and a provenance piece. The parameter check bit pc and mark check bit mc, which have been derived from the encoding parameters, can be used later in one-bit checking and during retrieval to determine with high probability if the correct provenance mark has been retrieved. Following are the formulas for the generation of pc and mc [9].

- $pc = hash(2, s@N_{pp})$
- $mc = hash(2, s@m)$

Where s is the amount of significant bits, N_{pp} is the number of distinct provenance pieces and m is the provenance mark. The formulas use @ for the concatenation of bit vectors [9].

5.4.2.3 Checking a Provenance Mark

Even though the embedding process produces an uncorrupted data set, it may not stay uncorrupted forever. The data set may get used, disseminated and even corrupted. Using one-bit watermarking on a data set, the method can determine, with a high probability, whether an embedded provenance mark is the original one or not (in which case the data has been modified) [9].
In order to find L_{sig} (the number of significant bits of a data point) and N_{pp}, the technique guesses the values for L_{sig} and N_{pp} and checks it using the following formula [9]:
$hash(2, d_0...d_{L_{sig}-1}@N_{pp}) = d_{L_{sig}}$
Using the parameter check bits, the guesses for L_{sig} and N_{pp} can be verified with high probability. This shows whether they are correct or not. In case they are correct and

provenance mark m:	0100110001001011
1st provenance piece pp^0:	01001100
2nd provenance piece pp^1:	11000100
3rd provenance piece pp^2:	01001011
4th provenance piece pp^3:	10110100

$$L_m = 16, N_{pp} = 4, L_{pp} = 8$$

Figure 5.9: Provenance pieces example [9].

$$suggest(\mathsf{d}, i) = \begin{cases} (\mathsf{pp}_j, j) & \text{if } j + k \mod L_m = i \text{ and} \\ & 0 \leq j < |\mathsf{pp}| \\ * & \text{otherwise} \end{cases}$$

Equation 5.10: Suggest bit i for data point d [9].

$$suggest(\mathcal{DS}, i) = \{suggest(\mathsf{d}, i) \mid \mathsf{d} \in \mathcal{DS} \wedge suggest(\mathsf{d}, i) \neq *\}.$$

Equation 5.11: Suggest for data sets [9].

bits $d_0 ... d_{L_{sig}}$ have not been corrupted, then $d_{L_{sig}}$ is the *parameter check bit* and is called *pc-consistent* [9].
Now suppose that we have a corrupted annotated data set DS for which the embedding parameter L_{sig} is known, and we have a guess at the provenance mark m. The method says that a corrupted data point d is *mc-consistent* with L_{sig} and m if:
$hash(2, d_0 ... d_{L_{sig}-1} @ m) = d_{L_{sig}+1}$
Using the mark check bit, which was added to each data point in the embedding process, the guesses for L_{sig} and m can be verified with high probability. If they are correct, and the bits $d_0 ... d_{L_{sig}+1}$ have not been corrupted, then $d_{L_{sig}+1}$ is the *mark check bit*. $d_{L_{sig}+1}$ is then called *mc-consistent*. This procedure is called *one-bit checking* [9].

5.4.2.4 Retrieving a Provenance Mark

To retrieve a provenance mark from a corrupted annotated data set, the method obtains provenance pieces from the data set and merges them to construct a best guess at the provenance mark. Even when the data set may have been corrupted, the provenance mark can still be obtained, as there is a lot of redundancy due to the way the provenance mark has been embedded. The process of retrieving a provenance mark is called *blind watermarking* [9].
The technique defines: $split(d) = (s, pc, mc, pp, k)$ as a function that takes a data point d, and based on L_{sig} and N_{pp}, returns information about the significant bits, check bits and provenance piece from d, where [9]:

- $s = d_0 ... d_{L_{sig}-1}$
- $pc = d_{L_{sig}}$
- $mc = d_{L_{sig}+1}$
- $pp = d_{L_{sig}+2} ... d_{|d|-1}$
- $k = hash(N_{pp}, s)$

Further the algorithm defines the suggestion for bit i of the provenance mark of a data point d as shown in Equation 5.10 [9].
Where $suggest(d, i)$ is either $*$ (no information found) or a pair (b, c) where b is the bit (0 or 1) that d suggests for the ith provenance mark and c is a natural number that stands for the confidence of that suggestions. The higher c, the less confidence it implies [9].
The method transfers the definition from data points to data sets: $suggest(DS, i)$ is a list of suggestions for bit i of the provenance mark, which are derived from the data points of data set DS. Formally this is represented in Equation 5.11 [9].
Now, given this formal definition, the method needs to compute the "best guess" for the ith bit of the provenance mark. One possibility is that the method selects the bit that the

$$allVote(L) = \text{round}\left(\frac{\sum_{(\mathsf{b},c)\in L} \mathsf{b}}{|L|}\right)$$

Equation 5.12: allVote [9].

$$\mathsf{m}_i = f(suggest(\mathcal{DS}, i)).$$

Equation 5.13: Guess method for m [9].

most of the suggestions propose, regardless of confidence. This method is called $allVote$ and is defined as shown in Equation 5.12 [9].
There also exist other methods, like choosing the bit that the majority of the most confident suggestions propose, or weighting the vote b from suggestion (b, c) based on the confidence c [9].
Finally, when given such a method like $allVote$, the best guess$m$ for the ith bit is computed as presented in Equation 5.13 [9].
With the final computed best guess m, the method can verify the provenance mark via one-bit checking, see Section 5.4.2.3 [9].

5.4.2.5 Evaluation and Analysis of Self-Identifying Sensor Data

Prototype tools have been developed for encoding provenance marks into data sets, which were used to performing one-bit and blind mark checking. Furthermore, tools for corruption of the data sets, such as rounding, truncating and sampling have been developed [9].
As reference values for the analysis, the following terms are used. The *pc-consistency score* is defined as the proportion of data points in a data set that are *pc-consistent* for guesses L_{sig} and N_{pp}. Furthermore the *mc-consistency score* is defined as the proportion of data points in a data set that are *mc-consistent* for guesses L_{sig} and m [9].
For mark checking, the analysis has shown that mark checking is efficient while under the effect of sampling, as the correct encoding parameters were able to get retrieved with high probability given a sufficiently large sample of an otherwise uncorrupted data set (at least 50 data points). It is also efficient under the effect of rounding, under the requirement that rounding to the nearest multiple k for $2 \leqslant k \leqslant 128$ was done, as *pc-consistency scores* were far above those expected for arbitrary parameters, allowing the successful retrieval of the encoding parameters [9].
Further the analysis has shown that the robustness of one-bit checking is very similar to that of recovering the encoding parameters, since one-bit checking uses just the mark check bit. This is why the same results for sampling and rounding were found here, similar to the parameter recovery mentioned before [9].
As for blind checking, the analysis has found that the effect of redundancy and sampling yields a high probability of blind mark recovery for larger samples. An interesting found was, that for the effect of truncation and rounding, the graph indicates that it "falls off a cliff": Blind checking does endure some amount of truncation very well, however, at some point, once too many bits have been truncated, the method is unable to recover the provenance mark. Figure 5.14 shows this. Higher redundancy yielded higher robustness to truncation. The experiments indicated that similar results can be found under the effect of rounding [9].

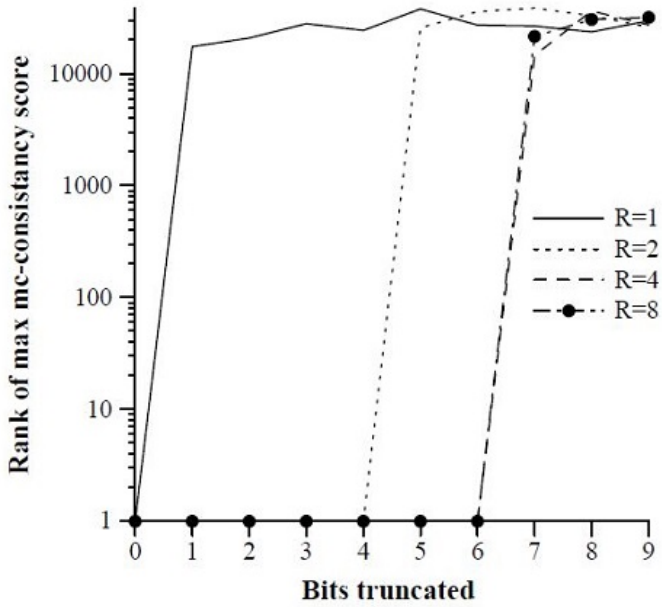


Figure 5.14: Rank of max mc-consistency score vs. truncation to nearest k, sample with 100 elements [9].

5.4.2.6 Conclusion of Self-Identifying Data

Self-Identifying data, which stands for sensor network data marked with its own meta data in an automatically recoverable procedure, is intended to support fair-use policies within the scientific community. Robust mark embedding, blind retrieval and one-bit checking algorithms are used to protect the data against transformations and modifications by users, which are expected that they perform on the data. These modifications encompass sampling, reordering, truncation and rounding. Evaluation and analysis has shown that the system is indeed very robust in various scenarios [9].

5.4.3 Copyright Protection in Wireless Sensor Networks by Watermarking

The authors of [10] propose a robust watermark algorithm for protecting the copyright of sensor data. They make use of the numeric characteristic of the sending time to embed the watermark. They also conduct experiments for performance comparison using 16 bit, 8 bit keys and none at all.

While developing this method the authors bore in mind that secure strategy is to be tightly connected with application. They also regarded that the equilibrium between the intensity of computation and security is ensured [10].

The embedding of the watermark is deliberately not done in the data. Altering sensing data in any way result in the data loosing value. Therefore, it was chosen to embed the watermark in the information concerning the transmission, more specifically the sending time. When transmitting information like sending time, source node identification, destination of the packet and length of data being transmitted is also included in a packet. It is only additional information and it does not have any specific connection to the perceptive sensing data, therefore the watermark is embedded in the sending time's least significant bit (LSB). Since it is done in the LSB the accuracy of the sending time is only affected in the minimal way.

The identification of the node is also used in the watermark, so that the actual copyright protection is made. One can identify the node by the watermark that has been embedded.

5.4.3.1 Assumptions for Using Watermarking as Copyright Protection in Wireless Sensor Networks

In order to use this method in a real world application the following assumptions must first be ensured [10]:

1. First assumption concerns the base station. It has to be robust and have enough computation power, so that it can deal with numerous packet information. The computational power is important so that the watermark extraction does not affect the lifetime of the whole network.

2. This method was developed for homogeneous sensor networks. This means that all sensor nodes of the network are the same in any way. They are then identified though an identification number not through their differences in hardware or software.

3. Since the method uses sending time for the embedding of the watermark, another assumption is that each node has the possibility to loosely synchronize time with the base station.

4. Last assumption concerns routing and transport protocols. For each node some kind of routing and transport protocols have to be carried out before transmission. This is not a concern in this paper, because the method does not influence routing and transport in such a way.

5.4.3.2 Watermark Embedding Algorithm in Copyright Protection by Watermarking

The watermark embedding process is done using the algorithm as it is described in Listing 5.15.
First the data that is to be sent is hashed to m_{data} and then the watermark is embedded bit by bit starting with the least significant bits. Each packet that is to be send receives one bit of the watermark.
The method uses sending time to embed the watermark, because for each packet the sending time is different. This makes the algorithm and watermark more robust. However, because the modification of the sending time is only performed on the least significant bit, we need several packets to transmit one watermark. Assuming the watermark is 8 bits long, then it would take 8 packets to transmit one watermark. The embedding of the watermark is done starting with the least significant bit to the most significant bit. The watermark is also encrypted with a secret key before it is embedded. This algorithms also enable the node to never add any more extra space to the packets that are to be sent.

5.4.3.3 Watermark Extraction Algorithm in Copyright Protection by Watermarking

The watermark extracting process is done using the algorithm in Listing 5.16.
First the position of the watermark is found, which is the sending time of the packet shown in line 3. Next the embedded bit is extracted, saved in wm. Once all bits of the watermark have been found, the watermark is decrypted. The decryption of the watermark, however, can only be done after all packets have been received that carry one watermark. By validating the watermark we can make sure the data that is send is from the right node. Even if a malicious node acts like part of the network it can be detected as such through the validation of the watermark.

Watermark Embedding Algorithm
Input: sensing packet P, watermark, ke;
Output: sensing packet P;
Steps:

1) read sensing packet P into data buffer buff
2) initialization, $P_{data} = P_{data} \& 0xff00$; $nu[8] = 0xfeff$; $lower = 0x01$; $cnt = -1$;
*3) read watermark into data buffer buff*1
4) $P'_{data} = key | P_{data}$
5) $m_{data} = hash(P_{data})$
6) if $i < 8$ *then*
7) $\quad m'_{data} = m_{data} \oplus (watermark >> i)$
8) else
9) $\quad i = 0$; *go to 3.*
10) end if
11) for $i = 0$ *to 8*
12) $\quad stime[i] = sent_time[i] \&\& nu[i]$
13) end for
14) $stime[0] = stime[0] || ((m'_{data} >> cnt) \& lower)$
15) $++cnt$;
16) $cnt = cnt\%8$

Listing 5.15: Algorithm for watermark embedding [10].

Watermark Extraction Algorithm
Input: sensing packet P
Output: watermark
Steps:

1) read sensing packet P into data buffer buff
2) initialization, $wm = 0$; $lower = 0x01$; $cnt = -1$; *go to 4.*
3) $stime[i] = sent_time[i] \& lower$
4) $wm = wm | (stime[0] << cnt$;
5) $++cnt$;
6) $cnt = cnt\%8$
7) if $wm \&\& cnt == 0$ *then*
8) $\quad$ *read out watermark*
9) end if

Listing 5.16: Algorithm for watermark extraction [10].

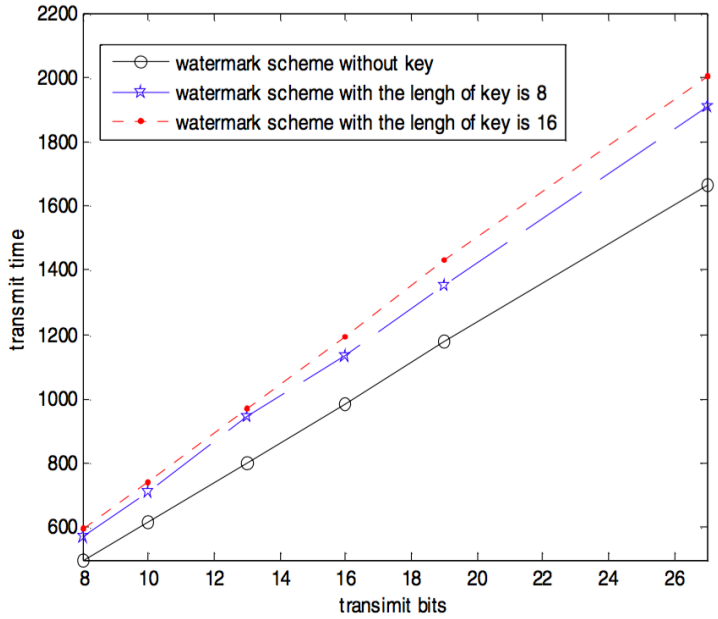


Figure 5.17: Proportion of data summarized or sampled in % [10].

5.4.3.4 Security Analysis of Copyright Protection by Watermarking

The authors of [10] only go about the security analysis in a theoretical manner. The goal of the method is to design a better secure scheme to achieve secure sampling and transmission and cut down the expenditure of network at the same time. However the authors aim at copyright protection, which means the secure hiding of the watermark, not so much the encryption of it.
They also use MD5 algorithm to hash values of the sensing data. So even if a malicious node captures a packet it is not able to make sense of the data at all, because the encryption of a MD5 hash value is only possible with great effort that is considered too big in most cases. Today MD5 is still used for encrypting passwords even though it has been shown that hacking is possible with much effort [10].

5.4.3.5 Experimental Results of Copyright Protection by Watermarking

To compare the suggested method not using an encryption, using a 8 bit and a 16 bit key a simulated network is used. The nodes are randomly distributed for the simulation.
The following conclusions were found from Figure 5.17: transmission rates of the watermark algorithms whose key length was 8 bits are 112%−115% times longer than the transmission rate of watermark algorithm without a key. A very similar conclusion can be drawn with the algorithm using a 18 bits key. The transmission rates are 120%−125% longer than the watermark with no key [10].

5.4.3.6 Conclusion on Copyright Protection by Watermarking

The method that has been introduced is able to protect the watermark of sensor packets. The authors were able to show this with a theoretical analysis and also using simulation. They also accomplished the goal of obtaining the trade off on computation power and cost.

5.4.4 Comparison of Methods for Authentication

The authentication method for self nodes based on watermarking in Wireless Sensor Networks is useful when sensors are deployed outside without being monitored by humans. Advantages of the method are that an adversary cannot forge data because he necessarily needs to know the shared parameters. Further, another positive property is that malicious data can be blocked and so error decisions can be prevented. The main advantage of the authentication method for self nodes is that the algorithm is very flexible and can be easily adapted to the requirements. A disadvantage of the method is the decreasing detection rate when the data is summarized. However, a larger value for the parameter v can counteract. Even though the storage is limited which acts as a restriction to the method.
The Self-Identifying Sensor Data method offers a solution to the problem, that the original source or *provenance* of wireless sensor data is often forgotten, even without malicious intent. Data sets generated using WSNs are of use for often not just the producer, but also subsequent generation of users. Its field of use is primarily in the scientific community and the public domain. The key benefit of the method is that it is very robust to benign and likely transformations, such as sampling, truncation, deletion/sampling, reordering and bit flips. The disadvantage of the method is that it is expected to operate under fair-use policies, it is not a security mechanism per se.
The copyright protection method using watermarks is only suitable for the actual copyright protection. The method is not suited to be used alone, but has to be applied together with other methods. However there are advantages to this method, like the fact that the watermark is not embedded into the data but rather into the sending time that. This way the value of the data is not decreased, as with other watermark embedding methods. The size of the packet that is to be transmitted stays the same, which poses a grave advantage. An additional benefit is that because the watermark uses the identification of the node, one can, after extracting and decrypting the watermark, distinguish from which node the packet has been sent. A disadvantage of the method is that it cannot be used alone as a security measure. It has to be combined, for example with secure routing. The assumptions that are made for this method to be able to work are also a disadvantage. As [10] shows in many experiments, the transmission rate is influenced when using this copyright protection method.

5.5 Summary and Conclusion

As we have shown, each and every method for authentication in WSNs has its advantages and drawbacks. The use of any method is always dependent on the application at hand. Nonetheless, there is always security to consider. Even though watermarking as such does not really relate to the ways the methods use the watermarks, it is still a security measure, meaning that the security of a method needs to be examined, *e.g.* by simulating attacks on one's own watermark. With watermarks different fields of research try their usage in implementing a more complex approach.
The difference between watermarking and identification is context dependent. A watermark can be used as identification. When attaching a watermark to one's documents, the reader of the document can clearly say to which company the document belongs. This could extend even as far as to individuals, not just organizations or groups, or in the case of WSNs to particular nodes.
However, watermarks primary function was never identification but verification. The purpose of watermarks was to be able to tell an original from a copy like with paper money. It was not its original purpose to link back to a company or individual but just to signal the value of the document for example. Ever since watermarking has been started

to be used in a digital form this separation of functioning has become harder. With digital watermarks we run into the difficult question of whether or not to extend the watermark that was to be used for verification to be now used as identification as well. In many cases there is only a small adjustment in the algorithms needed to achieve this new goal.
Regardless, such a distinction is not always desirable. When taking a look at the probability of all commercial content being watermarked, we would ask ourselves if watermarks are actually the best solution for the problem at hand. The point lies in knowing whether the purpose of watermarking content is to identify whom it belongs to, or if the purpose is only to signal that the copy that is being viewed is a legit one that was paid for. Here distinction is essential. If we only want to identify the author of the content a simple signature would suffice, but using a watermark would also be an option.
Watermarks are meant to be only detected by someone who knows that they are there. In contrary, a signature can be seen by anyone. This is also the distinction that has to be made concerning applications of these two concepts. Sometimes we do not want such a strong identification to be revealed to anyone and therefore hiding it using a watermark would be appropriate. What both methods need is protection from unwanted extraction and reuse as own. For this, other security precautions need to be taken.
Another optional feature of watermarks is concealing. People or individuals that know about the watermark could view it and use its provided verification, while others cannot. Now if every content would have a watermark by default it would change the situation. One would know that a watermark is embedded, but not visible. Still, to use the feature of verification one would have to extract the watermark.
Considering total traceability of content using watermarks in all content would lead to the situation which could be described as *"Big Brother is watching you"*, meaning that all content would be trackable. Economic players, such as music artists, would profit from this, as pirating and illegal copying of data will get harder. On the other hand, individuals would suffer from a loss of privacy. This leads to a trade off between economic benefit and privacy protection.
Besides, the total traceability of all content is not just imaginary, but also difficult to put into practice. Equipment manufacturers are not voluntarily willing to provide watermark detectors to their products, as this generates additional costs without benefits. A solution for this could be to force them to include watermark detectors by law. However, with enough resources one could always find a workaround and thus continue piracy.
The provided thoughts will lead to only more questions concerning this matter, making the discussion turn in a circle. For this reason we will leave it at this.
In conclusion, a discussion whether or not to use watermarks and why is endless. The answer is always dependent on the application at hand but also personal interest and knowledge. Therefore, we are not entitled to say that watermarking should be used for identification rather than a signature. In both cases the advantages and disadvantages are so different that a comparison is almost impossible.

Bibliography

[1] Biermann, Christopher J.: *Handbook of Pulping and Papermaking*; San Diego, California, USA: Academic Press. p. 171. ISBN 0-12-097362-6, second edition, 1996.

[2] Fernando Marincic, Loren Schwiebert: *Handbook of Sensor Networks: Algorithms and Architectures, Chapter 1 - Introduction to Wireless Sensor Networking*; John Wiley and Sons, Inc., New Jersey, USA, 2005.

[3] Xiaomei Dong, Xiaohua Li: *An Authentification Method for Self Nodes Based on Watermarking in Wireless Sensor Networks*; Northeastern University, Liaoning, China, 2009.

[4] Adrain Perring, John Stankovic, David Wagner: *Security in Wireless Sensor Nerworks*; Published in Communications of the ACM, June 2004. `http://cacm.acm.org/magazines/2004/6/6495-security-in-wireless-sensor-networks/fulltext`.

[5] Tanveer Zia, Albert Zomaya: *Security Issues in Wireless Sensor Networks*; International conference on Systems and Networks Communications, ISBN: 0-7695-2699-3, October 2006. `http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=4041555&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxpls%2Fabs_all.jsp%3Farnumber%3D4041555`

[6] All-Sakib Khan Pathan, Hyung-Woo Lee, Choong Seon Hong: *Security in Wireless Sensor Networks: Issues and Challenges*; ISBN 89-5519-129-4, DOI: 10.1109/ICACT.2006.206151, Februar 2006. `http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1625756&isnumber=34121`

[7] Antonio Galluzio, Romana Pernischova: *Trusted Computing in Wireless Sensor Networks*; Seminar: Communication Systems, University of Zurich, June 2015.

[8] Ingemar J.Cox, Matt L.Miller: *The first 50 years of electronic watermarking*; 4 Indepencdence Way Princeton, NJ 08540, April 19, 2002.

[9] Stephen Chong, Christian Skalka, Jeffrey A. Vaughan: *Self-Identifying Sensor Data*; ACM/IEEE International Conference on Information Processing in Sensor Networks, New York, USA, 2010. `http://dl.acm.org/citation.cfm?id=1791223`

[10] Rong Xiao, Xingming Sun, Ying Yang: *Copyright Protection in Wireless Sensor Networks by Watermarking*; International Conference on Intelligent Information Hiding and Multimedia Signal Processing, 2008.

[11] watermarking.sit.fraunhofer.de. `http://watermarking.sit.fraunhofer.de`, November, 2015.

[12] G. Warner-Allen, K. Lorincz, M. Welsh, O. Marcillo, J. Johnson, M. Ruiz and J. Lees: *Deploying a Wireless Sensor Network on an Active Vulcano*; IEEE Internet Computing, March, 2006.

[13] H. S. Ng, M. L. Sim and C. M. Tan: *Security issues of Wireless Sensor Networks in healthcare applicaations*; BT Technology Journal, April, 2006.

[14] Y. Qian, K. Lu, D. Tipper: *Towards Survivable and Secure Wireless Sensor Nerworks*; IEEE, 2007.

[15] Frank Hartung, Martin Kutter: *Multimedia Watermarking Techniques*; Proceedings of the IEEE, Vol. 87, No. 7, July 1999.

Chapter 6

Applying Social Network Analysis (SNA) to P2P File Sharing

Andreas Schaufelbühl, Robin Stohler, Benjamin Bürgisser

Social Network Analysis (SNA) is a widely used methodology to investigate relational networks. It is a method to gain insights into network structures, identifying important individuals in networks and provides findings about the network community. This report investigates the possibilities of the application of SNA to P2P Systems. Based on its popularity BitTorrent (BT) was chosen as a main example and the possibilities of modeling it as a graph are examined. Another part of this paper are network-centric as well as node-centric measurements which are interpreted for BT networks. On the ground of found interpretations and models the possibilities of optimizing the BT system is discussed.

Contents

6.1 Introduction **125**
6.2 Background **125**
6.2.1 Social Network Analysis 125
6.2.2 Peer-to-Peer File Sharing 126
6.3 Related Work **130**
6.4 SNA Measurements **130**
6.4.1 Network-centric measurements 130
6.4.2 Node-centric measurements 132
6.5 Different Approaches to Model the Network **136**
6.5.1 Different possibilities to model a network 136
6.5.2 First model of a BitTorrent Network 139
6.5.3 Second Model of a BitTorrent Network 142
6.6 Optimization **143**
6.6.1 Objectives and Problems 143
6.7 Summary and Conclusions **145**

6.1 Introduction

Even with video streaming platforms on the rise, peer-to-peer (P2P) applications still account for a large portion of the global Internet traffic. According to Sandvine [22] BitTorrent (BT) alone is responsible for 21% of the upstream and 6% of the downstream in Europe (not considering traffic from mobile devices). In Asia those numbers are even larger with 48% of the upstream and 19% of the downstream caused by BT. Analysis and optimization of the internal structures of large P2P networks therefore is subject to numerous studies. In [15] significant performance boosts were achieved using measurement-based optimization of such networks, which shows the potential of the field.
In this paper the possibility of applying Social Network Analysis (SNA) to P2P networks is discussed. SNA is typically used to study individuals or groups and their relationships within a social network, yet its main ideas may hold for other kinds of networks as well. As above mentioned numbers by Sandvine show, BT is the most popular P2P file sharing system. It was therefore selected as an example based on which the qualification of SNA for analyzing P2P networks is investigated. Another goal of this work is to find interpretations for the common SNA measures, which suit P2P-systems and their properties.
After a brief introduction to SNA and P2P networks in Section 6.2 an overview of related work is given, that influenced this paper. A deeper insight into possible SNA metrics is provided in Section 6.4. Based on the background information in the first part of this paper, different approaches to model a P2P network are discussed in Section 6.5 and possible optimizations are debated in Section 6.6.

6.2 Background

In this section the necessary background about the fields of SNA and P2P networks is provided, which are the foundation of the topics discussed in this paper.

6.2.1 Social Network Analysis

SNA is a method used to examine the structure within a network of individuals or groups. Such a network may be represented as a graph with nodes corresponding to the individuals or groups in question and edges illustrating a kind of relationship or interaction among them. As an example, in [4] SNA is applied to the social organizations of enterprises defining employees as nodes and their social interactions as edges. In [23] the use of SNA is discussed in the context of online social networks with users and groups as nodes of the graph. More applications of SNA are provided in Section 6.2.1.1.
Those examples show, that it is important to well-define the scope and parameters of the analysis, as the field of application of SNA is versatile. Parameters may include the type of network, the time frame for the analysis, or quality parameters [23]. After this initial step, data is gathered according to the defined parameters and scope. Data acquisition in the case of P2P networks is further discussed in Section 6.2.2.2.
Depending on the focus of the analysis the resulting graphs may be used to identify communication problems, inefficiencies, clustering and much more. The metrics to quantify the necessary properties for analysis are discussed in Chapter 6.4.

6.2.1.1 Applications of Social Network Analysis

First structural approaches to sociology were defined in the nineteenth century by Isidore Auguste Comte [9] and in the early 1930s sociometry was developed, which for the first time contained all the defining features of modern SNA. Since then the generality of the structural approach of SNA has been recognized and it has been applied in a wide range of

fields. In [17], network analytic articles published between 1984 and 1999 were examined and an almost linear growth in the number of areas in which SNA has been applied was found. As a result, SNA is not limited to social structures of human actors and can be applied to networks of machines (such as peers in BT) or even web pages in the PageRank algorithm [18].

Online Social Networks

In the form of Online Social Networks (OSN), virtual social structures have developed with users connecting and interacting online. They are expected to increasingly mirror real-life society and relationships and offer new possibilities for studying social structures on a large scale [5]. Moreover the data is clearly defined by constraints provided by the structure of the OSN. SNA is not only used by many studies yet there are even online tools available for the public to easily analyze their own online network. One of those tools is part of Wolfram Alpha [2]: After typing "Facebook report" into the search field, this functionality allows users to generate a report consisting of various charts and statistics about their friend network on Facebook.

Terrorist and Crime Networks

Both organized crime and terrorist organizations rely on social connections and their structure can be analyzed for investigative purposes [13]. This allows to identify important actors or links further to be monitored and aids identifying effective interventions to destabilize the network.

PageRank

In the late 90s, Google co-founder Lawrence Page introduced an algorithm to order search results which ranks web pages based on the graph of the web [18]. In this graph a web page corresponds to a node that may link to an arbitrary number of other web pages (outedges) and can itself be linked to (inedges). Highly linked pages are generally considered to be more important than pages with fewer links leading to them. Additionally the importance of the pages linking to a page is considered, as well as the number of the outgoing links of those pages. This concept is shown in Figure 6.1: Pages with no incoming links have the lowest PageRank (yellow circles). Pages with a lot of inedges have a higher PageRang (blue circle). If a very important page (green circle) links to only one other page (orange circle), it will have a high PageRank even if it has a low number of inedges.

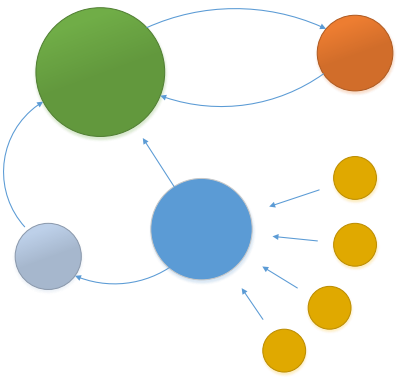


Figure 6.1: Graph representing web pages as circles, growing in size with their PageRank

6.2.2 Peer-to-Peer File Sharing

A P2P network is a self-organized system consisting of autonomous entities called peers, which share distributed resources. Every peer may act as a client and as a server simultaneously meaning it both supplies and consumes resources such as computing power or data within the network.

Over the years, several different P2P networks have emerged such as Napster, Gnutella or BT. Some of them use a completely decentralized approach, meaning that no central entities are used, as it is shown in Figure 6.2. Therefore any entity can be removed without loss of functionality. This technology is used in Freenet [16] and Gnutella 0.4 [11]. One of its main advantages is the lack of a single point of failure. However finding content in such a system is more complex than in a client/server-system as there is no central entity holding an index of all peers and their content. Instead flooding is used to search the network. In Gnutella 0.4 this means, that a peer broadcasts a request to all its neighboring peers which again forward it until the Time-to-Live (TTL) of the message is 0. The TTL usually starts with the number 7 and is then decremented every time the message is passed from one peer to another [16]. This leads to two scalability issues: On the one hand the TTL mechanism may terminate the search even if not all peers in the network have received the request. On the other hand flooding can result in high signaling traffic between peers and therefore a lot of bandwidth is used without content being exchanged.

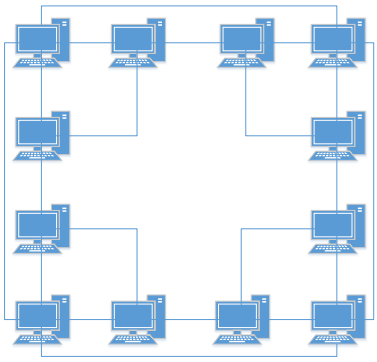


Figure 6.2: The peers in Freenet and Gnutella 0.4 without any central entities

To handle aforementioned problems with fully decentralized systems, some protocols are based on a hybrid approach that uses both central entities and P2P-technology. As an example Napster used a central server (Figure 6.3) where peers needed to register or log on to as they relied on its index to find content. If the desired data was available a connection between the requesting peer and a peer holding a copy of the data was established and at this point content could be exchanged without the central entity being involved. This meant that the central server was both the bottleneck of the system and its potential single point of failure.

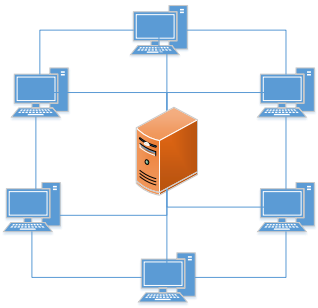


Figure 6.3: Napster's central look-up Server (orange) used to establish direct connections between peers (blue)

Another example for a hybrid system is Gnutella 0.6 that was designed to improve scalability compared to Gnutella 0.4. Peers with fast connections can be elected to become

ultrapeers [11], which are colored orange in Figure 6.4. Leaf-nodes in the network send their Query Routing Table (QRT) to the ultrapeers they are connected to. Therefore a request for a file does not have to be forwarded to the leaf nodes connected to an ultrapeer if none of them have the file stored.

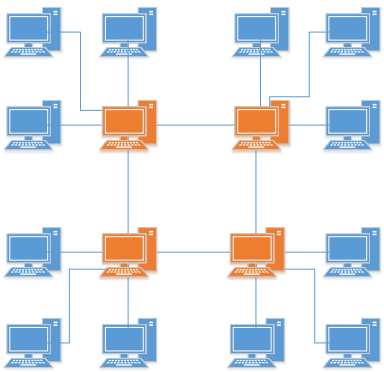


Figure 6.4: Leaf-nodes (blue) and ultrapeers (orange) in Gnutella 0.6

6.2.2.1 BitTorrent

In this paper BT serves as the main example of a P2P protocol. It can be categorized as a hybrid P2P-system as it uses so called trackers. Those are centralized servers that don't hold content themselves but track the swarm and gather some information about download and upload rate for statistics. The swarm consists of seeders and leechers. Seeders already have the complete file stored and therefore only upload data to leechers. Leechers have not yet completely downloaded the file and both upload parts of the file they already have and download parts they don't have. In Figure 6.5 an example of a swarm consisting of three leechers and three seeders is shown, which are all connected to a central tracker.

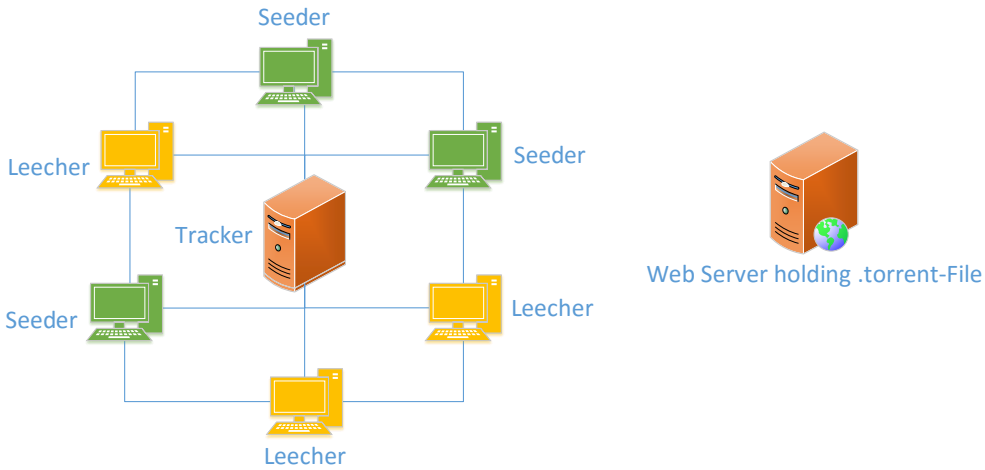


Figure 6.5: Seeders (green), leechers (yellow) and trackers in BT

To access a tracker and enter the swarm a peer needs to first download a .torrent-file from a web server, which is shown on the right side of Figure 6.5. This file contains the URL of one (or several) trackers and information about the file such as its length and name. The peer may then connect to the tracker and receive a random list of peers that have the requested file. According to [6] this is part of the standard tracker algorithm because of the good robustness properties of random graphs. As a next step the peer connects to

peers contained in the list. In BT a file is split into pieces of 256 KB by default which are again divided into sub-pieces of typically 16 KB. After connection has been established a peer reports which pieces it already has and requests pieces it does not have. This allows for downloading a file from multiple sources at once.
A peer may simultaneously download from as many peers as it is connected to and follows a strict policy for piece selection always requesting the rarest piece first. Using this technique rare pieces are replicated quickly within the network reducing the risk of them getting lost and increasing the download speed for other peers requesting the piece. As rare pieces are usually stored only on one or very few machines, when download starts an exception is made and instead of the rarest piece a random first piece is requested. This technique is used because it is important to get a complete piece as soon as possible to offer for download to other peers and therefore sub-pieces should be downloaded from multiple sources at once.
To maximize download rate BT uses a variant of tit-for-tat. This means, that a peer tries to upload to only a fixed number of peers which offer the highest download rate. Uploading in BT is usually called unchoking whereas choking is a temporary refusal to upload to a peer [6]. To evaluate which are the best connections BT uses a technique called optimistic unchoking. This implies that on a regular basis (usually every 30 seconds) a peer randomly chooses a peer to upload to and at the same time chokes the peer with the lowest download rate.
Sometimes a peer is chocked by all the peers it was downloading from and would have to wait for the next optimistic unchoke to find peers with higher download rate. For this reason a peer uses anti-snubbing. This means that a peer will choke a peer if it does not receive any pieces from it over a certain amount of time (usually one minute).
Shortly before finishing a download a peer enters endgame mode to prevent delays caused when requesting the last few pieces from peers with slow transfer rates. Instead in endgame mode a peer sends requests for all sub-pieces to all peers. To avert too much bandwidth from being wasted cancels are sent for arrived sub-pieces and the endgame period is kept short.
When a peer has finished downloading it turns from a leecher into a seeder, meaning it only uploads data without downloading. It then changes its uploading strategy to prefer peers which it has better upload rates to, as those are usually the peers that only a few (if any) other peers upload to.

6.2.2.2 Data Acquisition

Whereas in some P2P-systems (e.g. Napster) a central server establishes the connections between peers and therefore may be able to gather information about the structure of the network, gathering data for analysis purposes is not as straight forward in most of the aforementioned systems. In BT we may have some information about the network, because the tracker keeps a list of peers and some statistical data, yet the uploading and downloading processes are controlled by the peers themselves.
Hughes [12] therefore proposes three different approaches to monitoring the network. In **Network-level Monitoring** code is deployed at suitable points within a network to observe IP-level packets and identifying the traffic that matches known behavior of P2P systems. The other two methods are performed at the application level. In **Passive Application-level Monitoring** a modified peer is deployed within the network that logs the messages it routes. This approach is suitable for networks that require peers to route resource discovery and network maintenance messages, as it is the case in a fully decentralized P2P-system. In **Active Application-level Monitoring** on the other hand the modified peer actively attempts to connect and interrogate the network, to gather more data.

Since BT is a hybrid solution based on trackers, the above mentioned methods may be of limited use. In [24] a method specifically tailored to BT is described as follows: An instrumented client first downloads a .torrent-file and extracts the trackers. It then contacts the trackers to retrieve information about the swarms, including the peers contained within the swarms. As a last step the peers are contacted to perform a handshake and listen to sent messages.

6.3 Related Work

Wang et al. [27] investigate the SNA structures in P2P network, a statistical analysis on P2P social networks was done. Six different major peer social networks and original topologies were analyzed on the basis of SNA properties such as degree distribution, clustering coefficient average path length, betweenness. Major peers were discovered which are contributing and also sharing having symmetric connections to each other which is rare in such a network.
Upadrashta et al. [25] suggest a social semantic routing the idea is to build a social network according to querys with a positive result, the positive results get saved as a friends list fitting to the category of file the original node was looking for in future if a similar file is needed the friends list will get asked if they do not have the file they may ask their friends list. This work is only based on Gnutella and not on BT. But showed a case of optimization in a P2P system.
Pinthong and Liliakatsakun [19] analyzed the function of BT in depth and tried to enhance the performance of the BT network elaborating two other approaches to select possible Peers to download from which are only location driven the Piecepicker Localized Algorithm and the Tracker Localized Algorithm.
Kas et al. [14] compared two datasets where different numbers of nodes were selected taking into account either randomly chosen, high closeness, high betweenness, high out degree or high Eigenvector to send malicious data in the network then the number of affected nodes was measured. Eigenvector turned out to be in general a better indicator of effectiveness then the degree. This showed the influence of the measurements in such a network.
Feldman and Chuang [8] listed several possible strategies to over come free riding.

6.4 SNA Measurements

There are two kinds of measurements used to analyze a social network. Measures addressing the importance and topologial embedding of a individual node are called node-centric measurements. Whereas network-centric measurements can be used to capture systemic properties of complex networks. In the following, a set of different measurements will be discussed and explained in a simple graphical example.

6.4.1 Network-centric measurements

Network-centric measurements are aggregate measures that can be used to capture systemic properties of complex networks. In the following, metrics describing the resilience of a network against failing nodes or links and the efficiency distributing information are discussed.
Network size:
The most basic type of measurements can be defined in two ways. One is to count the number of all edges, the other one is to count all numbers of node. It is a simple

measurement, but it is also the most meaningless, because it only says how big a network is.Figure 6.6 shows a example of counting the network size

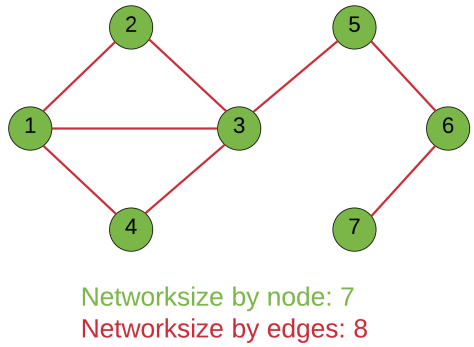


Figure 6.6: Example of Network size counting

Network compactness:
The compactness of a network is defined as the ratio of the number of edges over the number of edges that could possibly exists. It does not say anything about the size of a network, but about how tightly a network is connected overall. Figure 6.7 shows a example of calculating the network compactness

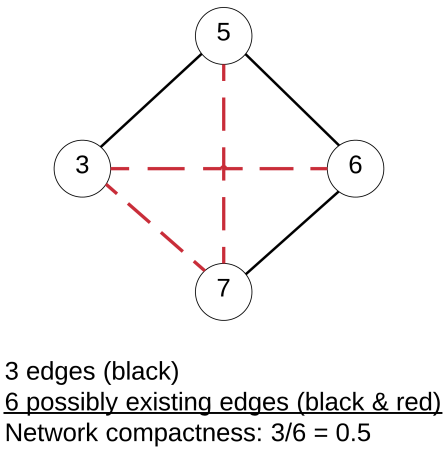


Figure 6.7: Example of Network compactness calculation

Average degree:
The average degree is the sum of all degrees divided by the number of nodes. It is a good measurement to compare a specific node with another one and identify if a node has higher or lower degree than the typical degree of the network has. Figure 6.8 shows a example of calculating the average degree.
Diameter and average distance:
The diameter is defined as the longest shortest path existing in a network between two nodes. Whereas the average distance is the sum of all shortest paths divided by the number of shortest path in a network. Having a high diameter, the network contains at least one really long direct path between two edges. Having a high average distance means that the typical path between any nodes of the network is long, what you generally want to avoid. Figure 6.9 shows a example of calculating the diameter.

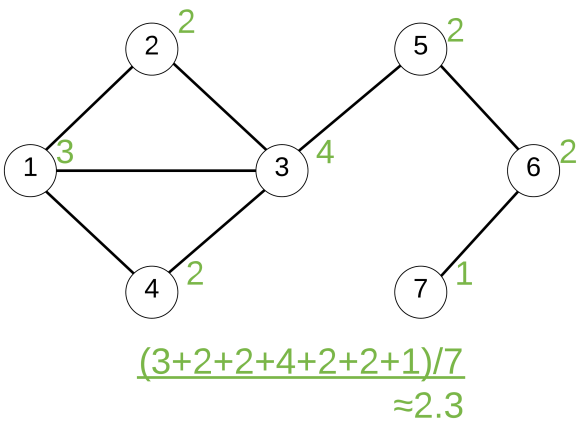


Figure 6.8: Example of Average degree calculation

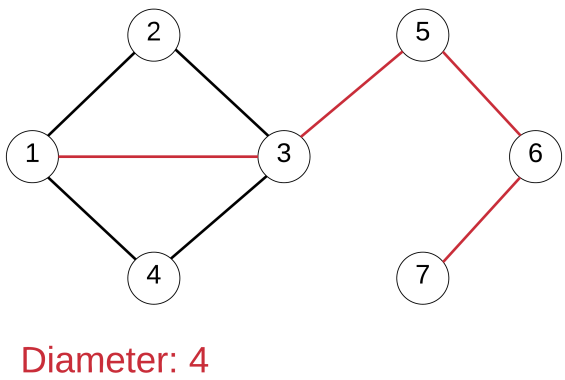


Figure 6.9: Example of diameter calculation

Measures of connectivity:
There are two measurements of connectivity: edge-connectivity and vertex-connectivity. Both describing the minimum amount of elements you have to remove of a network before it falls apart into multiple isolated sub-networks. The edge-connectivity of a connected graph describes the number of edges to remove, whereas the vertex-connectivity describes the number of vertices or nodes to remove, that the network falls apart. Figure 6.10 shows a example of calculating the connectivity of a network.
Global Clustering coefficient:
The Global Clustering Coefficient describes the ratio between triangles and triplets. Triangles are three neighboring nodes fully connected with each other, whereas triplets are three neighboring nodes, which are at least connected with two nodes. That means, that triangles are always triplets as well by definition. Figure 6.11 shows an example of calculating the Global Clustering coefficient of a small network.

6.4.2 Node-centric measurements

Node-centric measurements describe the importance or centrality of individual nodes as well as their embedding in the topology of a network.
Degree
The most basic node-centric measurement is the degree. It is defined by the number of edges connected to the observing node. It is a very simple measurement, but makes the

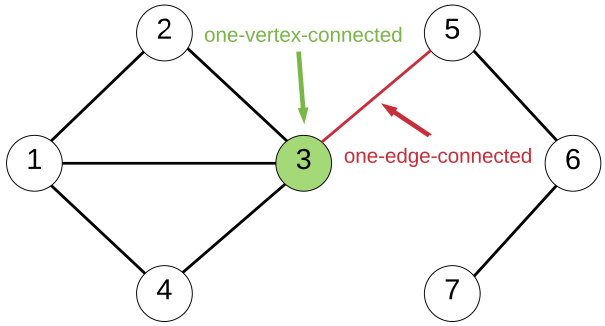


Figure 6.10: Example of connectivity calculation

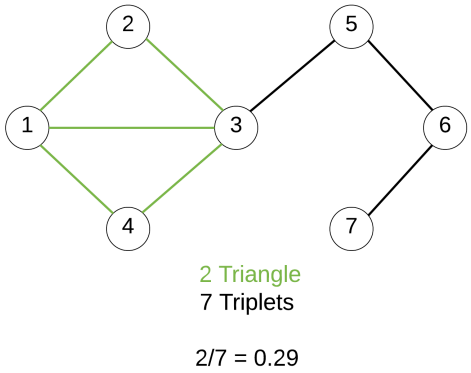


Figure 6.11: Example of Global Clustering Coefficient calculation

node comparable with each other. A node with high degree only says that it is connected with many other nodes, but there is no statement of the importance of a its connected edges. Figure 6.12 shows an example of define the degree of a node.

Betweenness centrality:

The Betweenness centrality is a more complex, but important measurement. The idea is to compare the shortest path going through a node with the shortest path not going through this node. Dividing the number of all shortest path between two nodes going through the third, observing node by the all the number of shortest path existing between the same two nodes gives you a value of importance for the observing node. The Betweenness centrality sums all this value of each combination together.

If a node has a high Betweenness centrality it is crucial for the network. It means that a lot of shortest path and therefore a lot of interaction goes through that node.

$$g(v) = \sum\nolimits_{s \neq v \neq t} \frac{\sigma_{\mathrm{st}}(v)}{\sigma_{\mathrm{st}}}$$

g(v): betweenness centrality

σ_{st}: total number of shortest paths between nodes s to t

σ_{st}(v): number of those shortest path between node s to t pass through v

Figure 6.13 shows the beginning of a Betweenness centrality calculation. There are two shortest path from node 5 to 1 and from node 6 to 1. All of them passing through node 3, but only one per pair passes through node 2. That is why node 3 get higher numbers

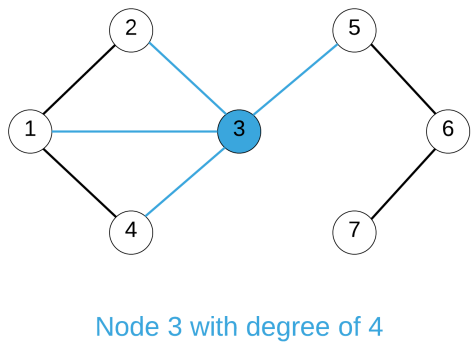


Figure 6.12: Example of node's degree definition

added to his Betweenness Centrality. To get the final Betweenness centralitiy of the node 3 or 1, all the combination of shortest path has to be checked.

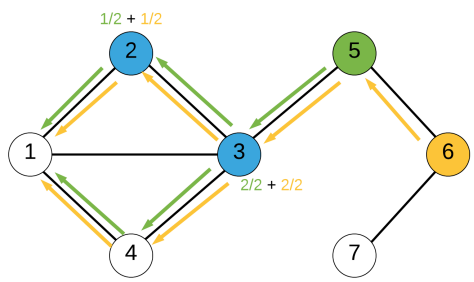


Figure 6.13: Example of node's Betweenness Centrality calculation

Closeness centrality:
Closeness Centrality describes the inverse of the farness of a node to all other nodes in a graph. The farness of a node is defined as the sum of its all shortest path to other nodes. Therefore, a central node will have a lower total distance to all other, but a high Closeness Centrality compare to a not-centralized node. It is a good measurement for showing central nodes. Downside of this measurement is that it only works with a network, where every component is connected to each other component, called connected graph. If there is an isolated cluster or node, this measurement can not be used because the distance between a node not connected to an other node is defined as infinite. Figure 6.14 shows a example of calculating the Closeness Centrality of the node 3.

$$C(x) = \frac{1}{\sum_y d(x,y)}$$

C(x): closeness centrality
d(y,x): shortest path between y and x
Eccentricity:
The eccentricity is related to the Closeness centrality, but instead of sum up all the shortest path, the eccentricity is the amount of the longest shortest path of a node.

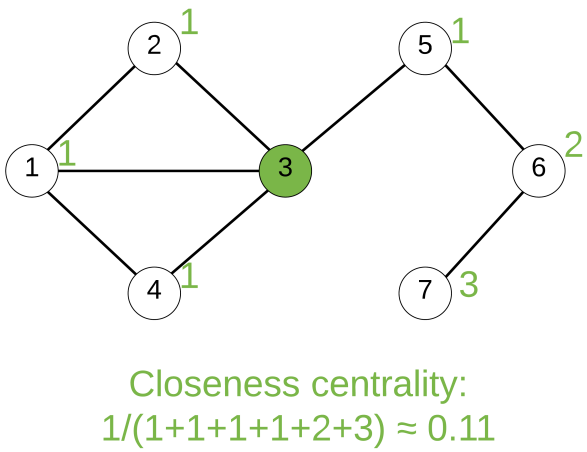


Figure 6.14: Example of Closeness centrality

Also the eccentricity measurement can only be used in a connected graph, otherwise the eccentricity if every node is infinite. The smaller the eccentricity of a node is, the more central it is and the greater the influence of the node is. Figure 6.15 shows a example of calculating the eccentricity of a network.

$$e(u) = \max d(y, x) : x \in V$$

e(u): eccentricity
d(y,x): shortest path between y and x

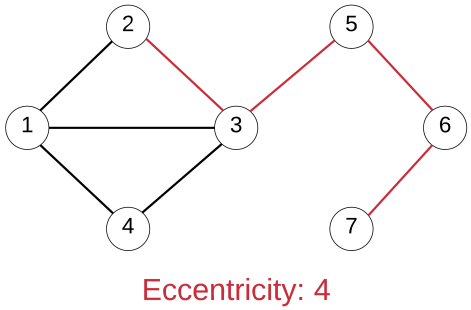


Figure 6.15: Example of eccentricity

Eigenvector centrality:
Eigenvector is a measurement of influence of a node in a network. It is a relative value with the idea that a node connected to high-scoring nodes contribute a higher eigentorvector of this node. The scoring unit can be other measurements like Degree, Betweenness Centrality and so on, based on the different algorithm. There are many different types of Eigenvector centrality measures. The most populars are the Google PageRank and Katz centrality, but there are many more. This paper does not explain these different measurements, because they are really complex and can fill a paper on its own.
Local Clustering coefficient
The local Clustering measurement does almost the same thing as the Global Clustering coefficient, described above. The only difference is that it focused on one node and not the entire network. It is a measurement to quantify how closely the neighbors of a node

are connected to each other. The neighborhood is defined as the immediately connected vertices to a node.

$$N_i = \{v_j : e_{ij} \in E \vee e_{ji} \in E\}$$

N_i: neighborhood of node v_i
The local clustering coefficient is then given by the proportion of the edges between all the neighborhood and the number of edges that could possibly exists. The number of possibly existing edges in a neighborhood of v_i with the number of vertices k_i is defined by:

$$\frac{k_i(k_i - 1)}{2}$$

Therefore, the local clustering coefficient for undirected graphs is defined as:

$$C_i = \frac{2|\{e_{jk} : v_j, vk \in N, e_{jk} \in E\}|}{k_i(k_i - 1)}$$

Figure 6.16 shows a calculation of the local clustering coefficent of node 3.

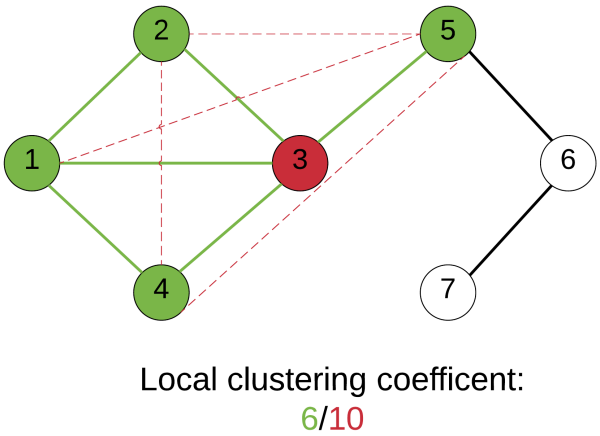


Figure 6.16: Calculation of the Local Clustering Coefficient of node 3

A high local clustering coefficient means, that the neighborhood of a node is very good interconnected and the node is within a cluster.
Coreness, k-core
The k-core of a network is the largest subgraph of connected nodes, where each node has at least the degree of k. Based on this definition, the coreness of a node is the maximium k-core to which it still belongs to. Figure 6.17 shows a example of calculating the coreness of a network.

6.5 Different Approaches to Model the Network

This section will discuss our approaches to construct a theoretical model of a Social Network with BT data and shows some construction possibilities.

6.5.1 Different possibilities to model a network

There are many different possibilities to construct a social network. Usually a social network maps a relation called edge onto two objects of the same class called nodes. These edges might have a direction, be weighted or it is decided that only edges are

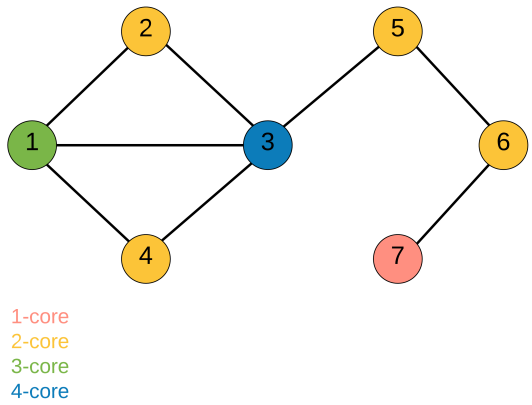


Figure 6.17: Example of coreness calculation

drawn in agrph at a specific point in time. Following some possibilities how to contruct a social network.

6.5.1.1 Time

Time is an important measurement of a graph. A graph is only valid in the measured time period. Therefore, every graph shows a snapshot of a well defined time interval. It is possible to indicate the growth of a network with different colors of nodes. Fig. 6.18 shows the number of nodes and the number of edges at each point in time. This shows that at certain time points the constructed graph is much bigger than in another point of time. This example indicates that old nodes and edges decay after a while without any action and reappear in the graph if some kind of action happens.

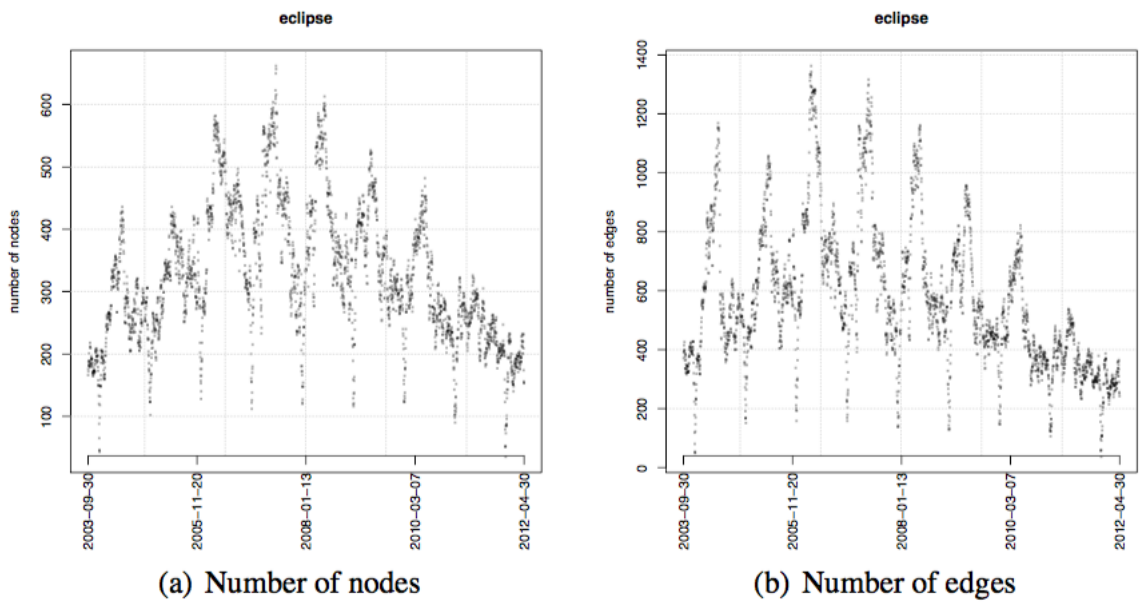


Figure 6.18: Shows the number of nodes and the numner of edges in a collaboration network changing over time [4].

6.5.1.2 Weighted

Edges might be weighted or unweighted. A weighted edge quantifies a measure of the connections between the two nodes. In terms of the P2P network a high weight would imply that multiple files have been sent on the same edge.

6.5.1.3 Directed

A graph might be directed or undirected. This depends if it is important to know who sent the message and who received it. An undirected graph treats every edge the same way and every edge will increase the degree whereas a directed graph is treating in-coming and out-coming edges differently having two types of degrees namely an in-degree and an out-degree.

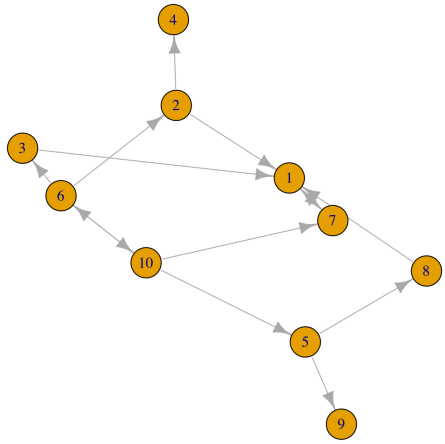


Figure 6.19: A directed network

6.5.1.4 One mode

A one mode social network is containing only one set of nodes that share the same type. One mode graphes are more important because most of the centrality measurements are only discussed for one mode graphs.

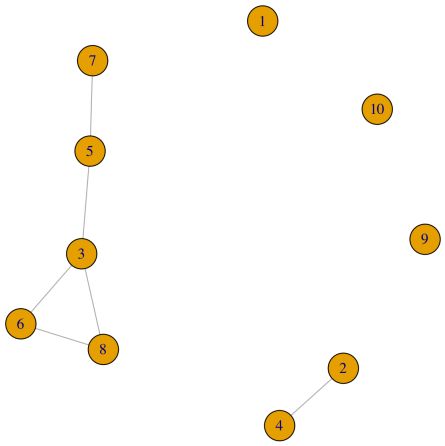


Figure 6.20: A one mode network

6.5.1.5 Two mode

A two mode Network contains two different types of nodes edges can only exist between these two different type of nodes. A well known example of a two mode Network is the Scientists collaboration network where the scientists are one type of nodes and the papers are another type of nodes. Two mode networks are usually projected into two one mode networks since as already discussed most of the SNA measures are only available on one mode networks.

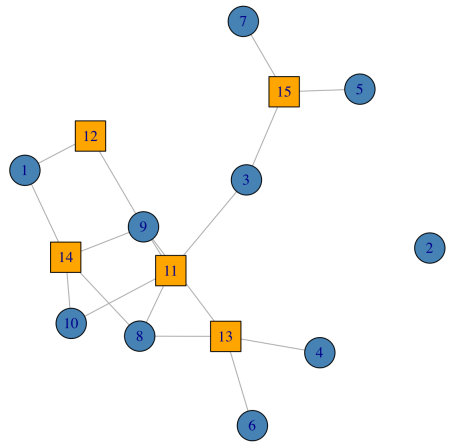


Figure 6.21: A two mode network

6.5.2 First model of a BitTorrent Network

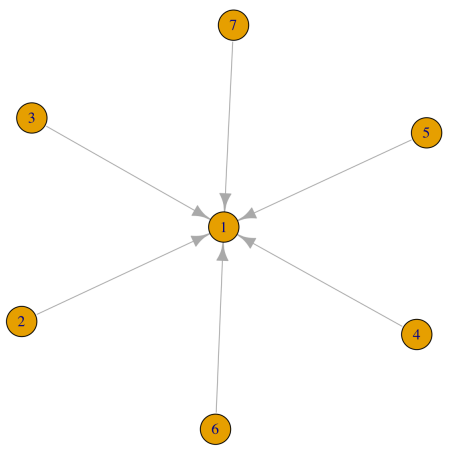


Figure 6.22: Shows the ego view of node 1 downloading from other nodes returned by the tracker. Such a model would result in a high degree while downloading a file. The in-degree of 1 would be 6 whereas the degree of the other nodes would be one.

A one mode graph would be built only with the users as nodes and with the chunks of files as the edges. A directed connection would be created from node A to node B if B was returned to A by tracker and B would send a chunk to A. Nodes with a high in-degree are leechers and nodes with a high out-degree are seeders. A momentary snapshot of this model, one peer downloading a file, would look like Fig. 6.22. This graph model would contain many smaller clusters connected with the peer that wishes to download files from the others. These clusters are built by the tracker who answers the request of the downloading peer with a finite number of peers that are known to hold the file. The selection of these peers happens randomly.This is shown in Fig. 6.23. In the end

if the download is completed the peer may keep the file in the BT application and stay connected to the tracker as a seeder or he might delete the file. It is known that only a small number of people contribute to P2P systems, where the rest of the nodes only consume [27]. According to this knowledge there exist a small number of nodes with high degrees these are the major nodes who might serve many nodes as source. These major nodes are some of the few nodes that have reciprocal connections to each other. Major nodes have consequently a high eigenvector [27].

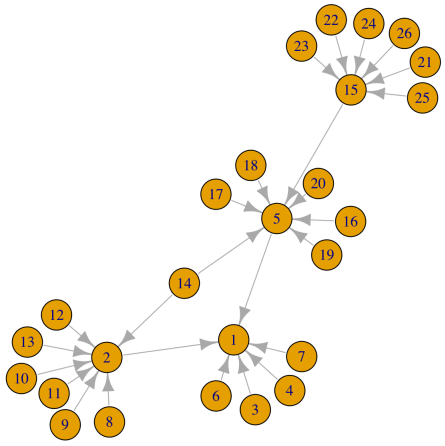


Figure 6.23: Shows the ego perspective of nodes 1, 2, 5, 15 downloading chunks from other nodes.

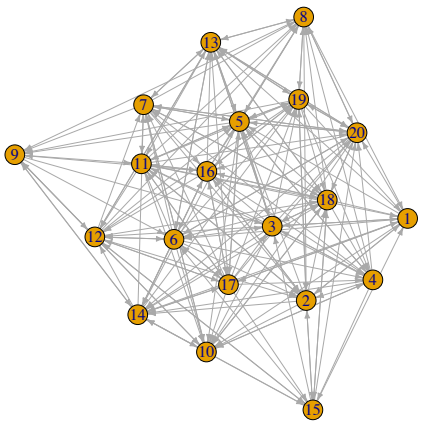


Figure 6.24: Shows the constructed graph with a mean degree of 20 and a mean in-degree and out-degree of 10 this could be a network of only one file and its seeder and leecher.

6.5.2.1 Issues with the model

This model has practical issues because it is constructed from the view of the downloading node. To receive this data every leeching node would have to report which node he received from the tracker, this would also ignore the process of the returned nodes by the tracker of downloading from each others.

6.5.2.2 Interpretation of the Measurements of the First Model

In the following, some of the measurements introduced in Section 6.4 are applied to the model.

Degree Centrality *How many different sources did this peer download files from? How many unique peers did this peer serve files to?*
Degree Centrality is in the BT network according to our model a measurement of the interaction between different nodes. A high total degree means that the Node received and sent many packages in the network. Where a high out-degree stands for a seeder and a high in-degree for leecher.

Eigenvector Centrality *Which peer is surrounded by the other, most popular peers?*
Nodes with high eigenvector centrality are very important. A high eigenvector means that an important node is linked or close to another crucial node. These nodes are the most effective nodes in spreading malicious content as shown in[14].

Closeness Centrality *How fast a new file will spread from this peer to the other peers in the network?*
Closeness is a difficult measurement to evaluate in a BT network, since the selection is randomly done by the tracker. In a usual social network the nodes might choose by them selves with whom they communicate or with whom they have a relation. This is not the case for BT in [26] it is said that closeness is correlating with degree, this might give us a hint. Because of a random selection of the node the distribution of the closeness might be similar to the one of a random graph.

Betweenness Centrality *If you can sniff packets, through which user confidential information is most likely to flow?*
With the betweenness it is a similar case with closeness.

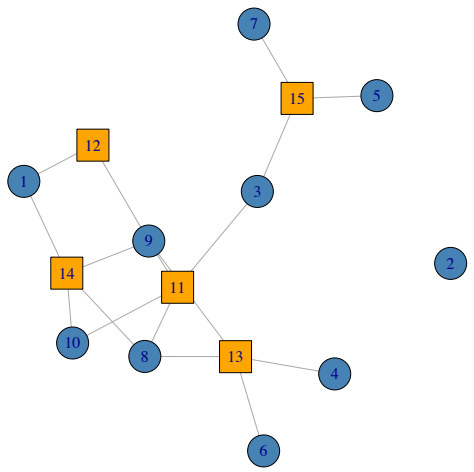


Figure 6.25: Bipartit or two mode graph

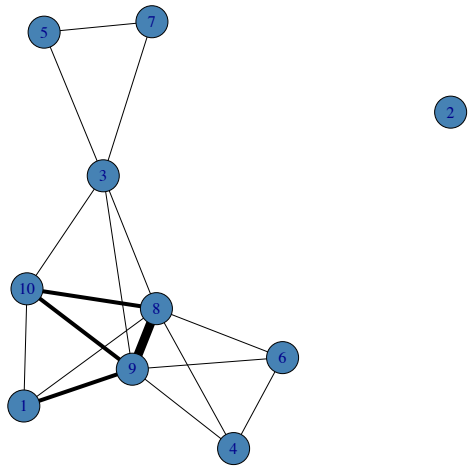


Figure 6.26: First projection

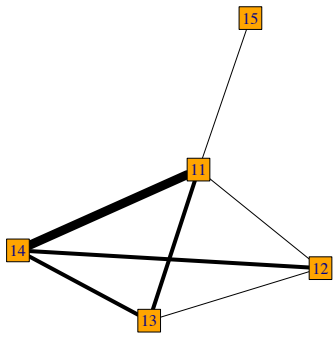


Figure 6.27: Second projection

6.5.3 Second Model of a BitTorrent Network

A more practical and less technical model could be constructed performing requests from a BT client to receive nodes from the tracker that have a specific torrent file. If this is done properly with sufficient torrent files a huge list with peers that provide the torrent files can be received. This is basically a two mode graph where peers are connected through torrent files as showed in Fig 6.25. To be able to calculate the SNA measurements we can project the graph we received into two one mode graphs. One that connects peers with weighted edges standing for how many torrent files these Peers have in common like Fig 6.26 and one graph that has torrent files as nodes and weighted edges as numbers of peers that offer the same file like Fig 6.27.
The construction of this second model returns a graph that links peers together which have the same torrent files in common this could be used as a recommendation system such as ”others also downloaded this torrent”. With this data profiles could be generated and such profiles of peers could be used by marketers to promote similar content what the peer already downloaded.

6.6 Optimization

The application of SNA has the goal to discover new knowledge from the examined subject. This can also show weaknesses like work overload of individuals or cut of points in the network. If these cut off points are removed from the network the network separates in to two parts. Depending on the field one might avoid such weaknesses and therefore, optimize de network building bridges or new connections to reduce the work load or data transmission load of one node. This optimization process might be split into several steps.

Steps of an optimization process[4]

1. Objectives of the optimization have to be clarified.
2. An adequate degree of abstraction has to be chosen to model the network.
3. Optimization
4. Qualify the network with a measurement to check wether or not the optimization was successful .

6.6.1 Objectives and Problems

Possible Objectives of an optimization process in the field of Application SNA to BT could be done taking into account following topics:

- Performance
- Availability
- System integrity
- Free riding
- Not completely decentralized
- Decentralized search
- Improve social collaboration

6.6.1.1 Performance

The question is in which step of the downloading process could a better selection of the nodes result in a better performance. The main performance enhancement in a P2P network like Gnutella could be reached if the selection of possible peers to download is not chosen randomly but deterministic. In this case [25] proposed for Gnutella the usage of a friends list categorized by the type of files this peer provided earlier. The only optimization application for the SNA might happen in step three where a random list is returned. A better performance could be achieved selecting nodes that are relatively closer than other nodes so it can be profited from a reduced transportation delay, and higher bandwidth because a local network[19].
The two algorithms proposed in [19] are:

Tracker Localized Algorithm To be able to run this algorithm the peer nodes have to be ordered by their autonomous system (AS) which might also belong to the same internet service provider (ISP). This ordering by AS can easily be done by the tracker according to the IP address.
The algorithm then tries instead of returning random nodes, which stands for great robustness [6], returning the closest nodes, preferably from the same AS like the requesting node. Because of the smaller distance and the same AS a higher performance may be reached.

Piecepicker Localized Algorithm This algorithm changes the Rarest first policy to the nearest first. The idea behind the rarest first is that as soon as the rarest piece is shared among multiple leechers the rarest is not the rarest anymore, this secures the diversity of the chunks in the network. Instead of the Rarest first the closes will be shared first this grants a better performance.

Friendslist approach Another approach would be if we would combine the friends list approach with the system of BT. Similar to the one used in [25], where nodes that already shared a file are saved as friends in a list categorized by the shared file. In future if a similar file is needed the query would be sent first to the friends that might have such a file. This approach works perfectly fine for Gnutella because queries are sent randomly to nodes, so if we preselect nodes with a higher chance to hold such a file we gain enhanced performance. The major issue in this approach is the application in the BT downloading process. We could change the policy of the tracker to only return nodes we already know, the chances that this would result in a better performance are very low but at least we could be sure that a node we already had contact can be trusted with a higher chance.

6.6.1.2 System Integrity

It is a problem that not every node might be trusted. Therefore a social network could help to avoid this issue. The application of the measures could provide a tool to evaluate someones trustfulness. Policies that favors only peers with a high degree and a high eigenvector would guarantee to get only chunks from very important and hence trusted nodes in the net.

6.6.1.3 Decentralization

BT is because of the tracker not fully decentralized. But this could be implemented. Nodes that have the same file build together a full connected social network. Well informed about all other nodes having the file. Instead of tracker information the torrent file would have information about possible nodes that have or had the file. Then if a node gets a request from an other node which desires to download the file, the first node can then select which nodes to return. If many returned nodes are not part of the network anymore each other returned node might return other nodes in his network. In this way the BT network would be decentralized. But this is again very similar to the Gnutella 0.4 approach and would therefore, use way more messages between these node groups.

6.6.1.4 Free riding

Free riding is the act of consuming and not contributing.
In P2P systems, where cooperation may cause significant communication and computation costs, rational thinking users may refuse to contribute their fair share of resources. Free riding should be automatically handled by the tit-for-tat policy, which evaluates the

download rate of leechers and monitor if a seeder uploads or not. Another policy that should work against free riding is anti-snubbing, when a peer realizes that after some time no pieces were exchanged from a neighboring peer, the leecher assumes it is snubbed by the other peer and stops uploading to it by the usual unchoking [19]. Even with such policies free riding might still be very successful.
In a SNA free riding nodes are easily identified with their out-degree which is below the median. This might also be the case for newcomers which just joined the network. That is why also other measures should be included to know whether or not someone is free riding.

6.7 Summary and Conclusions

First an overview is given on the multiple branches of the topics SNA, BT and Gnutella. Then the different measures of SNA were discussed and possibilities of how to build a network are shown. A network of the BT network was constructed. Possible optimization with the application of SNA were thought of and finally the interpretation of some measurements were described. The BT network is not the best network to apply the SNA model since everything is well defined and already very sophisticated. In particular the tracker returns a certain number of nodes which are already known to have the file. This random selection determines how the graph looks like but this is also a point where policy changes could include the metrics of SNA. Thus Free riding might be a big issue but a major loss of performance is not measurable.

Bibliography

[1] Miltiadis Allamanis, Salvatore Scellato, Cecilia Mascolo: *Evolution of a Location-based Online Social Network: Analysis and Models*, Proceedings of the 2012 ACM conference on Internet measurement conference, November 2012, pp. 145-158.

[2] Wolfram Alpha LLC: *Wolfram Alpha*, December, 2015, `http://www.wolframalpha.com/`.

[3] Tao Ban, et al.: *Practical network traffic analysis in P2P environment*, Wireless Communications and Mobile Computing Conference (IWCMC), July 2011, pp. 1801-1807.

[4] Valentin Burger, et al.: *Social Network Analysis in the Enterprise: Challenges and Opportunities*, Socioinformatics - The Social Impact of Interactions between Humans and IT, Springer International Publishing, August 2014, pp. 95-120.

[5] Salvatore A. Catanese, et al.: *Crawling Facebook for social network analysis purposes.*, Proceedings of the International Conference on Web Intelligence, Mining and Semantics (WIMS '11), 2011.

[6] Bram Cohen: *Incentives Build Robustness in BitTorrent*, Workshop on Economics of Peer-to-Peer systems (Vol. 6), May 2003, pp. 68-72.

[7] Andrew Fast, David Jensen, Brian Neil Levine: *Creating Social Networks to Improve Peer-to-Peer Networking*, Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining, August 2005, pp. 568-573.

[8] Michal Feldman, John Chuang: *Overcoming free-riding behavior in peer-to-peer systems.* SIGecom Exch. (Vol. 5, No. 4), pp. 41-50, July 2005.

[9] Linton Freeman: *The development of social network analysis. A Study in the Sociology of Science.*, Empirical Press, 2004, `www.researchgate.net/profile/Linton_Freeman/publication/239228599_The_Development_of_Social_Network_Analysis/links/54415c650cf2e6f0c0f616a8.pdf`.

[10] Tobias Hoßfeld, et al.: *On the computation of entropy production in stationary social networks*, Social Network Analysis and Mining 4.1, December 2014, pp. 1-19.

[11] Daniel Hughes, Geoff Coulson, James Walkerdine: *Free Riding on Gnutella Revisited: The Bell Tolls?*, Distributed Systems Online (Vol. 6), June 2005.

[12] Danny Hughes, James Walkerdine, Kevin Lee: *Monitoring Challenges and Approaches for P2P File-Sharing Systems*, International Conference on Internet Surveillance and Protection, 26-28 August 2006.

[13] Renee C. van der Hulst: *Introduction to Social Network Analysis (SNA) as an investigative tool*, Trends in Organized Crime 12 (No. 2), 2009.

[14] M. Kas, L.R. Carley, and K.M. Carley: *Monitoring social centrality for peer-to-peer network protection.* Communications Magazine, IEEE, pp. 155-161, December 2013.

[15] T. S. Eugene Ng, et al.: *Measurement-based optimization techniques for bandwidth-demanding peer-to-peer systems*, INFOCOM 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications. IEEE Societies (Vol. 3), 2003, pp. 2199-2209.

[16] Andy Oram: *Peer-to-Peer: Harnessing the Power of Disruptive Technologies*, 2001, `http://dl.acm.org/citation.cfm?id=558412`.

[17] Evelien Otte, Ronald Rousseau: *Social network analysis: a powerful strategy, also for the information sciences.*, Journal of information Science 28 (No. 6), 2002, pp. 441-453.

[18] Lawrence Page, Sergey Brin, Rajeev Motwani, Terry Winograd: *The PageRank Citation Ranking: Bringing Order to the Web.*, Technical Report, Stanford InfoLab, 1999.

[19] N. Pinthong and W. Lilakiatsakun: *Performance of bittorrent-like p2p file sharing systems inspired by ALTO.* TENCON 2013 - 2013 IEEE Region 10 Conference (31194), pp. 1-5, Oct 2013.

[20] Dongyu Qiu, Rayadurgam Srikant: *Modeling and Performance Analysis of BitTorrent-Like Peer-to-Peer Networks*, ACM SIGCOMM Computer Communication Review (Vol. 34, No. 4), September 2004, pp. 367-378.

[21] Mengyin Ren, et al.: *P2P networks monitoring based on the social network analysis and the topological potential*, 2014 IEEE Conference on Communications and Network Security (CNS), October 2014, pp. 492-493.

[22] Sandvine Incorporated ULC: *Global Internet Phenomena*, December 2015, `https://www.sandvine.com/trends/global-internet-phenomena/`.

[23] Atul Srivastava, Anuradha, Dimple Juneja Gupta: *Social Network Analysis: Hardly easy*, 2014 International Conference on Optimization, Reliabilty, and Information Technology (ICROIT), 6-8 February 2014, pp. 128-135.

[24] Camilo Andres Gonzalez Toro, Tom Chothia: *Analysis of BitTorrent Peers' Behavior and Monitoring Trends*, University of Birmingham, 2009, `www.kaspersky.com/images/camilo_andr\%D1\%83s_gonzalez_toro.pdf`.

[25] Y. Upadrashta, J. Vassileva, and W. Grassmann: *Social networks in peer-to-peer systems.*, HICSS '05. Proceedings of the 38th Annual Hawaii International Conference on System Sciences, 2005, pp. 200, January 2005.

[26] Thomas W. Valente, et al.: *How Correlated Are Network Centrality Measures?*. Connections (Toronto, Ont) (Vol. 28, No. 1), pp. 16-26, 2008.

[27] Fang Wang, Yamir Moreno, and Yaoru Sun: *Structure of peer-to-peer social networks.*, Phys. Rev. E, 73:036123, March 2006, `http://arxiv.org/pdf/physics/0603222.pdf`.

[28] Fatmawati Zifa Zaukani: *Using Social Network Information to Improve the Performance of P2P Networks*, Masters Dissertation, (College of Engineering, Seoul National University), February 2008, `http://my-groups.de/altmann/images/Publications/Theses/fatma_thesis.pdf`.

[29] Yipeng Zhou, Tom Z. J. Fu, Dah Ming Chiu: *Statistical modeling and analysis of P2P replication to support VoD service*, INFOCOM, 2011 Proceedings IEEE, April 2011, pp. 945-953.

Chapter 7

How to Get Users to Use CoinBlesk

Cyrus Einsele, Mervin Cheok, Patrick Widmer

This seminar thesis presents Bitcoin technology and mobile payment solutions that rely on this technology. With the title being "How to get Users to Use CoinBlesk" the seminar thesis provides answers to this question. The seminar thesis is structured in a bottom-up fashion, first explaining the foundations of the underlying technology and then following up with different applications. The same approach was chosen to answer the question at hand: breaking up the problem into smaller ones, analyzing each independently, and providing solutions. Finally, advantages and disadvantages of different solutions are discussed. As a conclusion the authors have suggested two different paths to take: establish the application within a niche group of users and/or implementing some sort of achievement system.

Contents

7.1 Introduction and Problem Statement 151
7.2 Related Work 151
7.2.1 Bitcoin 151
7.2.2 Mobile Payment Services 154
7.3 Approaches 160
7.3.1 B2C 160
7.3.2 C2C 161
7.4 Solutions 162
7.4.1 B2C 162
7.4.2 C2C 163
7.5 Summary and Conclusion 163

7.1 Introduction and Problem Statement

Even though this seminar thesis is called "How to Get Users to Use CoinBlesk", the Bitcoin system will be introduced, since CoinBlesk is built on top of the Bitcoin system.
In other articles a distinction is made between the lowercase and uppercase version of the word "Bitcoin". In this report this convention will be continued and the system will be refered to as "Bitcoin" and the currency as "bitcoin".
Bitcoin is a payment system that relies on peer-to-peer technology and operates without a central authority. Furthermore it is open source software. A brief technical overview will be given in a following chapter.
The focus of this seminar thesis lies on the CoinBlesk application, and more specifically how to increase it's usage. CoinBlesk is a mobile Bitcoin payment solution developed at the University of Zurich, Switzerland. CoinBlesk is currently in its second iteration. Whereas the first version was a centralized payment system, the second version is now more decentralized. The system consists of an Android application and a Java server. To address the question - How to increase CoinBlesk usage - this talk also examines a part of the mobile payment systems landscape in Switzerland.

7.2 Related Work

First, an overview of Bitcoin will be presented. In a second part, the focus lies on mobile payment systems. CoinBlesk will be analyzed, as well as alternatives, such as Paymit or Tapit.

7.2.1 Bitcoin

First, the Bitcoin system will be described in a few sentences. Next, an overview of the history will be presented, including the development of Bitcoin, as well as alternatives. After that the technical background will be analyzed. The scope of this analysis is kept to a minimum. It contains elements such as the Bitcoin protocol or the block chain as well as aspects of security of such a system. In the end of this subsection current trends will be presented, followed by a summary.

7.2.1.1 Description

Bitcoin is a payment system that is based on peer-to-peer technology. One advantage of networks based on peer-to-peer is, they have no need for a central authority. Transactions are processed collectively by the participants of the network.

7.2.1.2 History

Bitcoin is a payment network and it is open source, meaning that everybody has access to the source code. Everybody can participate in the developing process [2]. In contrast to traditional currencies, that are controlled by national banks, Bitcoin does not need a central authority. Bitcoin doesn't belong to anyone [2].
Bitcoin was first mentioned in a white paper called "Bitcoin: A Peer-to-Peer Electronic Cash system" in 2008 [3]. One year later Bitcoin was released as open source software by Satoshi Nakamoto. Up to date it is still argued whether or not this pseudonym belongs to a single person or a group of people [4].
Currently, bitcoin is still the most popular crypto currency [1]. However, it is by far not the only crypto currency. Today, there are already over one hundred crypto currencies. The field is growing very fast. One can easily create its own crypto currency and it will

probably continue to become even simpler to do so. Two of the better known alternatives to bitcoins are ripple and litecoin [1] [5].
Ripple is a virtual currency, but also an exchange platform for all kinds of currencies. The list of goods that can be exchanged include crypto currencies as well as traditional currencies and even credit in form of mobile phone minutes [5].
Litecoin is another crypto currency and payment system. One difference is the interval in which transactions are processed. While it takes Bitcoin at least ten minutes for the processing of a transaction the same action takes only two and a half minute in Litecoin [5].

7.2.1.3 Technical Background

Every time a computational step is finished (i.e. a new block is added to the block chain), a reward in form of bitcoins is distributed to the pool that has successfully computed this block. This process is called mining [6]. The reward is a predefined amount of bitcoins. In the beginning it was fifty units. Every four years this amount is halved. At this moment, it is twenty-five bitcoins [15].
Already at the beginning it was clear that the total amount of bitcoins has to be restricted to a specific amount. This is a controversial point that is frequently discussed among members of the Bitcoin community. It is getting more important due to the fact that the amount of bitcoins in circulation comes closer to the initially defined limit [1]. The maximum amount of bitcoins to be mined was set to twenty-one million units [7]. One party of the community would like to increase the limit, while another more conservative group excludes such a possibility, because such an action would probably affect the price of a bitcoin.
The fact that the amount of bitcoins is finite is one fundamental difference to traditional currencies, where national banks theoretically are able to create an infinite amount of their currency. They would never do that, because as a direct consequence the value of the underlying currency would fall to zero. But it is an important tool when considering inflation.
The only condition in order for a user to participate in the system is the operation of a client application, which is compatible to the rules of the Bitcoin protocol. This allows computers with different underlying components to participate in the same system. An alternative would be to use an online service, that provides this functionality. The latter is especially important when considering devices with low computational capacity, such as smartphones or other mobile devices. In these cases, an alternative would be to use an SPV client, that uses a method for verifying transactions without downloading the whole block chain. Thus, mobile devices are enabled to make use of Bitcoin.
A Bitcoin client is an application that has to be installed on a computer in order to use Bitcoin. There are different type of Bitcoin clients. There are Bitcoin clients that download the whole block chain, others only download a fraction of the block chain. The block chain is stored decentralized and redundant at any point in time. In case of transactions the block chain is updated and synchronized over the network.
Payments within Bitcoin are handled using pseudonym addresses. Bitcoin is able to produce multiple addresses for each individual user. The system doesn't provide identification of the participating entities. Neither, it guarantees total anonymity. The reason for this lies once again in the design of the system. Every transaction with the corresponding addresses is stored and thus publicly available within the block chain to prevent double-spending.
The block chain is a decentralized database, that is used by Bitcoin for controlling the transactions. One reason to implement Bitcoin was that traditional payment systems need a lot of trust by the users [8].

In order to corrupt the block chain and thus hit the heart of Bitcoin, a user, or a pool of users, would need to overcome fifty percentage of the computational power of the whole system (i.e. all contributing nodes together). At this point this task is almost impossible for a single user to achieve. However, participants in the network form pools to increase their chance to get a reward [6]. If an attacker would overcome this mark he could manipulate the system. An attacker would have the possibilities to invert previously confirmed transactions or to not accept any new transactions. However, since the user themselves have an interest in maintaining the system, they would not want such a scenario to happen [9]. This is due to the fact, that bitcoins would loose their value. It has been shown that pools that have overcome this boundary, take actions to reduce their computational capacity relative to the whole system, either by splitting up the group or similar mechanisms [10].
Further aspects of security include the prevention of multiple spending of the same bitcoin as well as the authorizations of payments such that they can be done only by the owner of a bitcoin. The latter is ensured with the use of private keys.

7.2.1.4 Current Trends

More than five years after the introduction of Bitcoin, the stock price is still volatile as shown in Figure 7.1. When looking at the development of the stock price per bitcoin over the last three years, one can see, that in October 2013, the price has suddenly started to rise heavily. Before this moment, the price per bitcoin was around one hundred euros. In November 2013, the price reached a maximum of more than one thousand euros per bitcoin. After that, the price has fallen and risen in an alternating manner. In this period the price fluctuated mostly between two hundred and five hundred euros. Compared to traditional currencies, this fluctuation is considered high.

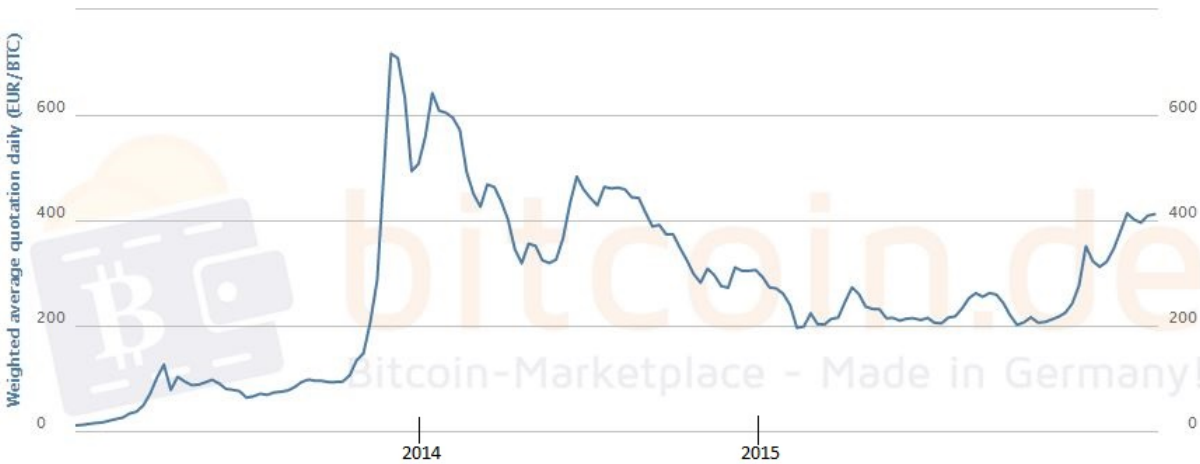


Figure 7.1: Development of the bitcoin price in euro [11]

As described in a previous section, users of Bitcoin are using pseudo addresses. Thus, users cannot be identified directly. This means, transactions cannot be mapped to individual users, since each user can have multiple pseudo addresses. In order to analyze the number of Bitcoin users, a different strategy has to be applied. A measurement could be the number of MyWallet users. MyWallet is an application for both desktop computers and mobile devices (e.g. smartphones), that allows users to make transactions [12]. As shown in Figure 7.2, the number of MyWallet users has risen heavily since January 2013. Up to date, November 2015, it has reached almost five million users.

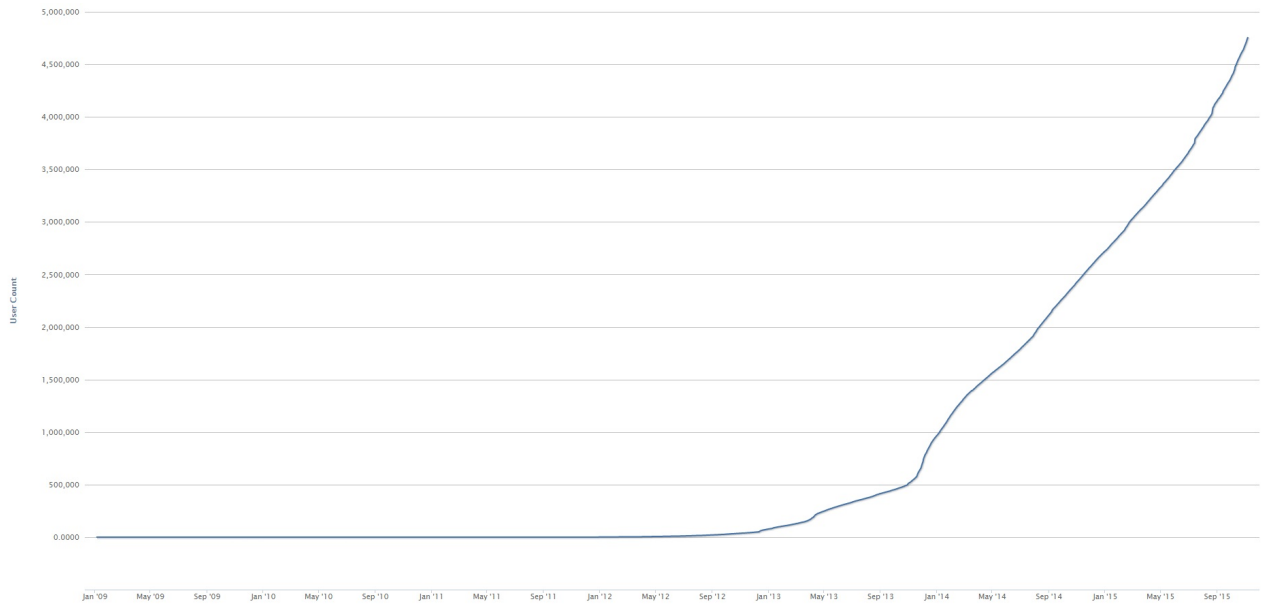


Figure 7.2: Number of MyWallet users [13]

Another measurement, is the number of transactions per day. How many bitcoin transactions are processed per day. As shown in Figure 7.3, the number of transactions is growing since mid 2012. Up to date, November 2015, around 150'000 transactions per day are done.

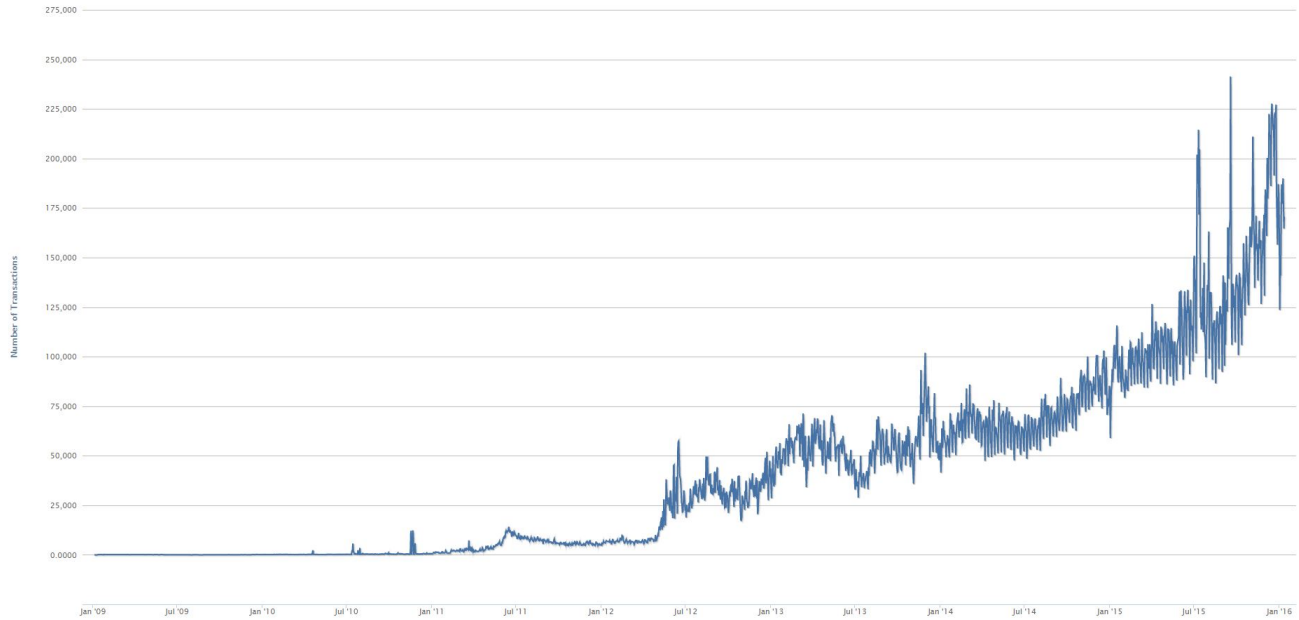


Figure 7.3: Number of transactions per day [14]

In order to summarize, the stock price of the bitcoin currency is still highly volatile, more than five years after its introduction. Time will show if it is going to stabilize with broader acceptance and wider use. The other two figures, both show a growing use of the Bitcoin system over the last two years.

7.2.2 Mobile Payment Services

This chapter discusses mobile payment services in general/app. Afterwards a selection of available mobile payment apps will be introduced. Main focus is the app CoinBlesk with a target marked in Switzerland. Accordingly other mobile payment apps, like Paymit and

Tapit which are also in the Swiss market and helps to identify patterns and commonalities for the introduction of successful mobile payment solutions, are chosen. Additionally a very successful mobile payment app from Denmark is considered.

7.2.2.1 Mobile Payment Services in Theory

This subchapter gives a brief introduction about the theoretical definition of mobile payment and tries to forecast in which direction this field is developing.

Definition

A mobile payment involves the transfer of money from one party to another [23] using their mobile device and the belonging services. This includes money transactions via mobile phones, chip cards, mobile wallets and other mobile devices (for example tablets, PDAs and notebooks), as well as the peer-to-peer payment platforms used with mobile phones [36].

Mobile payment apps typically are downloaded to a smartphone, and will be set up with a bank account or attached to a credit/debit card. Often the phone number is used as identifier and a PIN code will authorize the payment [23]. A newer flavour is the use of Bitcoins instead or in addition of a bank account or credit card.

The mobile payment landscape is using several technologies like near field communication (NFC), QR Codes and/or SMS. Various stakeholders like financial institutions, mobile network provider and regulators are involved that make the mobile payment landscape highly complex [37].

Forecast

Growth forecasts for mobile payment services have been very positive. The development of digital communication and transaction channels in the banking environment is becoming a key role. In the near future the integration of mobile devices will be promoted, since in banking these devices gain a more important role. Not just in direct communication with the financial service provider, but for the customer side transactions business as well [36]. Despite these encouraging forecasts, however, the reality looks quite different, and the situation is often disappointing for those firms offering mobile payment services [22].

"When it comes to mobile payment, the chicken-or-egg analogy is frequently used to describe the challenge facing merchant and consumer adoption issues. On the one hand, merchants are unwilling to invest in the systems needed to enable a mobile payment transaction unless there is consumer demand. On the other hand, consumers will not use mobile payment systems unless merchants accept them" [37].

7.2.2.2 Profile Construction Plan

Every app will have a short introduction with focus on the app's background story, a slight overview of the special features and on the unique selling proposition or additional value (claimed by the app itself).

To build a basis for comparison for the selection of mobile payment apps, it is needed to put them in a standardised form. Therefore the information and specifications will be categorised and presented in a tabular form. The content/categories will be:

- **Producer/Provider** of the app.
- **Lifespan** - Release date and possible end date.
- **Origin** - Country where the app was developed.
- **Distribution geographically** - In which countries the app is used?
- **Distribution numbers of users** - Specifies an estimate of the number of users of the app.

- **Payment service type** - What kind of payment technology is used by the app (for example Creditcard, Debitcard, Bitcoins, SMS, and so on)?
- **Supported platforms** - On which mobile platforms is the app available?
- **Initial effort/requirements** - What requirements have to be fulfilled to use the app? What effort has to be done by the user, so that he can use the app (for example, does the user have to register to a certain service, does he need a special sim card or does he need an account from a certain bank)?
- **Fees** - Will the user be charged for using some functionality of the app?
- **Currencies** - What currencies are supported by the app?

To gain this data a document analysis will be performed. The data sources include the web appearance of the app and published articles of scientific origin, technology magazines or newspapers.
The execution of the document analysis contains multiple iterations. The first iteration is used to find and identify possible data material. The second iteration has the goal to screen the material and get an overview. The second iteration results in a first idea for the necessary categories. Every iteration goes more into detail until a sufficient data basis could be gained.
It is possible that the categories have to be adjusted during the document analysis.

7.2.2.3 CoinBlesk

`https://bitcoin.csg.uzh.ch`
As mentioned previously, CoinBlesk is a student project of the department of informatics at the University of Zurich. CoinBlesk is an open source mobile online wallet that allows to exchange Bitcoins payments between its users. This is achieved by two-way near field communication (NFC) transactions. Release date was 01.09.2014. CoinBlesk has been used in a isolated test environment at the University of Zurich in the canteen ZFV Mensa Binzmuhle in Oerlikon for a short time period [24].
Functionality
CoinBlesk offers private user to private user transactions as well as private user to business transaction functionality. A CoinBlesk terminal can be built with only a few components [24].
The use of NFC allows one participant of the transaction to be offline (at least one party must be connected with the Internet). This allows the use of CoinBlesk in areas with unstable network coverage [24].
Registration Process/Setup
Users can get the CoinBlesk app in the Google play store. He then has to register a new account and verify his email address, and setup his account. Then the user can prepay Bitcoins into his CoinBlesk app.
A summary about CoinBlesk is shown in Table 7.1.

7.2.2.4 MobilePay

`www.mobilepay.dk`
In 2013 the Danske Bank released the MobilePay service. The aim of the Danske Bank, was to provide a solution to address the need for very simple money transfer. The idea of Danske Bank was to make money transfers between people and businesses as simple as sending a text message [29].

Table 7.1: CoinBlesk survey in tabular form

CoinBlesk	
Producer/Provider	University of Zurich
Lifespan	2014 - dato
Origin	Switzerland
Distribution	
- geographically	Zurich, Switzerland
- number of users	< 100 (estimation)
Payment service type	Bitcoins
Supported platforms	Android 4.4 and above, needs NFC support
Initial effort / Requirements	User needs Bitcoins
Fees	According to Bitcoin
Currencies	Bitcoin to CHF

MobilePay found great acceptance from the very beginning. At the first day, the MobilePay app was downloaded 25'000 times. The app counted 500'000 users after four months and two years after start it counted over 2 million users. Today, MobilePay is offered in Finland and Norway as well. MobilePay is accepted in more than 14'000 shops and per year there are over 120 million transactions done with MobilePay [29].
MobilePay does not charge fees for private users, for business users individual fees will be charged [28].
Functionality
MobilePay offers a simple way to send or request money from other users (the mobile number of the other party is everything that is needed). MobilePay is also accepted by businesses and shops and even online shops. MobilePay offers the possibility for other apps to use it as integrated payment service. Charities and NGOs accept donations via MobilePay [28]. MobilePay also offers a bonus program.
Registration Process/Setup
A private user can use MobilePay by downloading the app to his mobile device and follow the registration process. Name, credit card or bank account information have to be declared. After that a security pin has to be entered. This simple registration is all that is needed for a private user to begin using MobilePay. Neither credit card nor bank account have to be from the Danske Bank. About two thirds of users are not Danske Bank customers [28]. In order to use MobilePay the mobile phone has to be connected to the Internet.
A summary about MobilePay is shown in Table 7.2.

7.2.2.5 Paymit

`www.paymit.com`
Paymit was launched in Mai 2015 by SIX, UBS and Zurcher Kantonalbank (ZKB) as the new peer-to-peer (P2P) payment app. Paymit allows private persons to send and request smaller monetary amounts to and from other private persons in a simple comfortable and secure way. The three providers claim that Paymit is the first mobile payment application based on a open bank standard. This guarantees that Paymit can be used from every person resident in Switzerland and independently to which bank they belong [32].
Paymit started with just the customer to customer functionality, but is planned to extend customer to business functionality (i.e. Point-of-Sale, In-App und E- Commerce) soon [32].
In the first week after release date, Paymit has been downloaded and activated over 10'000 times. Mainly the UBS version, followed by ZKB and than SIX [33].

Table 7.2: MobilePay by Danske Bank survey in tabluarform

MobilePay	
Producer/Provider	Danske Banke, Trifork
Lifespan	2013 - dato
Origin	Denmark
Distribution	
- geographically	Denmark, Finland, Norway
- number of users	> 2 million
Payment service type	Bank account, credit card
Supported platforms	iOS, Android, Windowsphone
Initial effort / Requirements	App installed, bank account or credit card, Internet connection
Fees	Free for private users,
Currencies	Norwegian Kroner (NOK)

The providers admit, that they have taken the Danish success story MobilePay as role model for Paymit [30].

Functionality

At the current moment of this report, the Paymit app supports only transactions between private users. Business functionality like with MobilePay are already planned [32].

Registration Process/Setup

A private user has to download the Paymit app on to his device and follow the registration process. The user needs a Swiss mobile number and a credit/debit card or a bank account. After the registration the user is set to go [31]. In order to use the Paymit app the mobile phone has to be connected to the Internet [31].

A summary about Paymit is shown in Table 7.3.

Table 7.3: Paymit Survey in Tabluarform

Paymit	
Producer/Provider	SIX, UBS, ZKB
Lifespan	2015 - dato
Origin	Switzerland
Distribution	
- geographically	Switzerland
- number of users	> 20'000 (estimation)
Payment service type	Credit cards, bank accounts
Supported platforms	iOS, Android
Initial effort / Requirements	mobile number, credit card or bank account
Fees	Depending on the Paymit provider SIX, UBS(free), ZKB(only free with certain ZKB accounts, otherwise 20 CHF per year and 2CHF per transaction), and so on
Currencies	CHF

7.2.2.6 TapIt

`www.tapit.ch`

In 2014 the Swiss telecommunication company Swisscom has launched Tapit, a digital wallet for the mobile phone. Swisscom had the idea of an online wallet that should be able to transfer money, collect bonus points and should function as an electronic key for doors (replacing common badges). At the beginning Tapit supported only the money transfer feature between private users whit a sim card from Swisscom and a credit card from Cornercard [26]. Support for other telecom providers and credit cards followed. The possibility to use a bank account could not be realised due to lacking support of the banks [27].
In the first year Tapit was downloaded 10'000 times. This did not meet the expectations of Swisscom, which lead them to the decision to stop the Tapit project and join Paymit an other mobile payment solution in Switzerland. The Tapit services are going offline in 2016. As one reason for this negative development Swisscom names the lack of collaboration with the banks and the missing possibility to link Tapit with a common bank account [27], which restrains a simple and comfortable use of the app.
Functionality
Tapit offers the possibility to transfer money to other private user and businesses. The Near Field Communication (NFC) technology allows a simple and quick contactless payment process. The use of NFC allows the user to pay without being online (note that at leas one party involved in the payment process must be online though). The payment process can be executed in under two seconds. Payments with an amount of 40 CHF and lower can be paid without pin confirmation. Payments above 40 CHF force a pin confirmation. Tapit is accepted by Migros and Coop the two biggest Supermarkets in Switzerland. The bonus point feature can also be used with Coop [26].
Registration Process/Setup
To use Tapit a special NFC compatible sim card is needed. A private user then can download the Tapit app from the Google play store and follow the registration instructions. Furthermore must the user link his creadit card with the Tapit app. Tapit supports only the mobile Cornercard prepaid/credit cards or the mobile Viseca master card. Usually the user has to order these special card separately. Furthermore the user has to link his credit card as well as to define a pin and personal settings, to pay with Tapit [25].
Due to the necessity of NFC Apples iPhone devices are not yet supported by Tapit. Since Apple has not yet released its NFC API [35].
A summary about Tapit is shown in Table 7.4.

Table 7.4: Tapit by Swisscom survey in tabluarform

Tapit	
Producer/Provider	Swisscom, Tapit
Lifespan	2014 - 2016
Origin	Switzerland
Distribution	
- geographically	Switzerland
- number of users	$<$ 50'000 (estimation)
Payment service type	mobile credit card
Supported platforms	Android
Initial effort / Requirements	App installed, special sim card, special credit card
Fees	Free for private users
Currencies	CHF

7.2.2.7 Using Cash in Denmark and Switzerland

The government of Denmark is trying to push digitalisation. Copenhagen wants to free smaller shops and businesses from using cash. In the future, volunteering businesses do not have to accept cash anymore [30]. In 2015 nearly 40 percent of Denmarks population is using MobilePay. This trend is recognisable across Scandinavia. In this region cash is used for less than six percent of all payments made [34].
In Switzerland there is a recognisable trend for an increase in cash less payments. In the year 2010 87 percent of the payments have been done with cash. Five years later cash payments make 71 percent of all payments in Switzerland and in 2020 there will be an estimated 60 percent of the payments done by cash [33]. Therefore payment solutions like Maestro card, credit cards or mobile/wearable payment will gain even more popularity.
In both countries (i.e. Denmark and Switzerland) the potential for mobile payment is growing. Although a clear distinction has to be made. The 71 percent of payments done with cash this year, shows that cash is still very common and popular in Switzerland. In Denmark on the other hand, cash is only used for 6 percent of all the payments. The change from cash to a digital payment method is nearly completed in Denmark and promoted by the government. Whereas Switzerland is still at the beginning of this change and the government has not yet shown any interest to be part of this development.

7.3 Approaches

Before thinking about how to change the current state (i.e. increase the amount of users) it would be beneficial to asses the current situation beforehand. When doing an analysis it is often good practice to break up the subject in to smaller, more manageable parts. Two transaction scenarios are likely to be prevalent: Business-to-customer (B2C) and Customer-to-customer (C2C). The Government-to-customer (G2C) scenario is omitted, because it can be assumed that the state will only accept its own currency for taxes and such. In the following paragraph the abbreviations given will be used to denominate each scenario.

7.3.1 B2C

It seems that B2C could potentially benefit a lot from such a application. The reasoning being, the design and the technology (namely NFC) used for this product caters to quick and comfortable transactions, which is something that is usually valued by consumers and businesses alike.
There was only one temporary use of CoinBlesk in a B2C environment (at Mensa UZH Binzmuehle), so there is not a lot of feedback from users available. Therefore this analysis is mostly going to be based on assumptions, and where available, data from related fields. In recent years retailers in Switzerland have tried to push systems that replace the traditional cashier and instead let the customers scan their items themselves [16]. Often (specifically in the case of the large retailer Migros) this system has been paired with cashless payment. Presumably the reasoning behind this practice is the fact that cash accepting machines tend to be bulkier and more expensive than a small credit card terminal. Along with this strategy retailers have started to implement Near Field Communication (NFC) technology, which is also used by CoinBlesk, into their registers. The fact that both major retailers in the country are pursuing this strategy does make a strong statement about the assumed needs of their customers, which includes almost any person in Switzerland. What we can gather from this is, that there is a market for contact-less payment systems such as CoinBlesk.

This begs the question, what keeps people from using contact-less payment systems that rely on Bitcoin. It can be argued, that currently there is no added benefit in using such a system for neither the businesses or the consumers. On the contrary: For the regular consumer, there are actually drawbacks to using a Bitcoin reliant system. They need to acquire some amount of this currency. And looking at the business side, there are actually drawbacks here as well. One very important part of a business is their accounting. Though it is not unusual for a business to work with multiple currencies, it can be assumed that it would present a major challenge to incorporate bitcoins into their existing accounting, because of the volatility [17] of the currency in comparison to currencies that are backed by national banks. This is not referring to Zimbabwean Dollars, which are maybe a bit worse in that regard, but to the major currencies that are used by companies in this country.
Another major hinderance that can be identified, is people's fear of the unknown. Or in this case their unfamiliarity with the Bitcoin system. At the time of writing there was no data from Switzerland available, but a online survey performed on behalf of the German news site "Focus.de" [18], has revealed the following information from consumers in Germany, regarding the Bitcoin technology: 51.1% of the participants have not heard the term "Bitcoin" before, and of the other rough half that has heard the term before, only 55.8 % knew that it was a digital currency. 12.2% of the participants guessed that "Bitcoin" is a measurement unit of data transmission speed, 8.9% guessed that it was play money for slot machines and another 4.3% thought that it would be a cordless screwdriver.
So to summarize there are three areas, that are currently holding CoinBlesk and similar applications back in a B2C context:

- Lack of (perceived) added benefit for either businesses or consumers
- Volatility of the bitcoin currency
- Bitcoin not being known to a wide population and the resulting lack of trust therein

7.3.2 C2C

Before talking about the current state of C2C relationship in regards to using Bitcoin-based applications, it should be clarified what exactly is in the scope of this analysis: The transfer of money between two humans. So for example if you go out to eat with a friend, and your friend doesn't carry cash, then you opt to pay for the meal, and they will pay you back using their phone and a application such as CoinBlesk. In the following paragraphs the factors that have been identified in the B2C sections are analyzed again, but in a C2C context. It can also be assumed that monetary transactions in a C2C context fall into one of these categories: Transaction between people that know each other in real life and transactions related to the sale of goods or provision of a service. The second category of transactions is potentially arranged through a middleman such as eBay or Ricardo.
In a C2C context, volatility is still an issue, but it's not as influential as it might be in a B2C context. While consumers are presumably also interested in storing their money in a stable currency, they typically tend to have less money in their accounts than a typical business would. This is not true for very wealthy people, but it can also be safely assumed that the majority of potential users don't fall into this category.
Since money is a sensible topic for many people, the problem of Bitcoin technology not being known among a wide population is likely still a problem in the C2C scenario. No one would open an account at a bank that they've never heard of before and therefore don't trust. Even people that have heard of Bitcoin technology before, might be reluctant to use it themselves, because they are worried about the whereabouts of their hard-earned money.

Now onto the aspect, that can be thought of as the key aspect in the C2C context: The (perceived) added benefit from using such a system. This begs the question, why would specifically this be the key aspect to evaluate? Because this is where a system such as CoinBlesk really shines. If both parties in a C2C transaction already have everything configured beforehand, actually performing the transaction is quite convenient, when compared to conventional methods. But as described in Section 7.2.2.1 there are other payment services, that allow for convenient transfer of money using mobile phones. So why should consumers opt to use yet another application, that requires extra effort to set up (acquiring funds in bitcoins)? Some answers to this question can be found in Section 7.4.2.

7.4 Solutions

In this section, solutions, that address the problems identified in the previous section, are going to be presented and their viability as well as their potential pitfalls are going to be evaluated. In addition to that, it is going to be discussed what methods other providers have used to promote their services and whether or not these methods could be applied to popularize a service such as CoinBlesk. However, in contrast to most such promotional programs one party in this particular scenario is missing: A clear beneficiary, that would profit from putting in the considerable effort (i.e. spending money and using other resources) necessary to make such a application popular. With other payment systems (or services in general), there is usually one party that initiates and promotes the system, because they plan to make money with it. This fact should be kept in mind, when trying to analyze other paths that have been taken to achieve a similar goal.

7.4.1 B2C

With thousands of transactions being processed in a B2C environment within every passing hour, one begins to wonder how to direct some portion of those onto CoinBlesk or a similar system. By nature of the product, this particular application fits for transactions that consumers want to perform quickly and effortlessly. So perhaps consumers wouldn't want to use CoinBlesk for example to pay their car or a big household appliance. This observation leads back to the retailers "self-checkout" programs mentioned in Section 7.3.1, that seem to pair well with payment systems like CoinBlesk, that are performed through the use of mobile phones. A commercial service would probably want to seek out a partnership with one or more retailers. This approach is probably not viable though in this instance.

The problem that consumers don't have a perceived added benefit that comes with using CoinBlesk or another Bitcoin-reliant system, is a likely good starting point. Privacy-conscious consumers could potentially be convinced, that their identity is protected better when using a Bitcoin-reliant system compared to conventional methods of payment. Security-conscious consumers could be informed about the independence of Bitcoins regarding national banks or government institutions in general. The aforementioned efforts could be undertaken without significant financial means. But there seems to be a small catch: these consumers present a niche within the general population. While it is not known how the actual quantitative distribution of consumers regarding their different sensitivities looks like, it is most likely safe to assume that a major part of the population does not share these sensitivities.

Concluding from the observations made in the previous paragraph, one way to promote CoinBlesk would be to provide a small benefit for users in this early phase. This benefit could be of monetary nature, but they don't need to be. Humans like to collect things for

various reasons [19]. So one possible form of benefit could involve collectible stickers that are acquired through each transaction, this could even be digital - so within the actual application. Another form would work like a money back program, that is known from credit card companies. Even if the percentage of money that is paid back is very small, this scheme might attract some users. Such a promotional program should only last a limited amount of time, and the expectancy is, that a portion of the users attracted would also continue to use the application, even without added benefit in the future.
Another approach could be to seek out businesses, whose customers are likely to be sensitive about privacy and/or security and try to convince them to establish a payment method that uses CoinBlesk. This could potentially result in a win-win situation, where the business is giving their customers more payment options (perhaps even the preferred method of payment for some customers) and CoinBlesk gains new users. Similarly to the last approach, there is the hope, that some customers will continue to use the application, also for other purposes. If the business, that incorporates CoinBlesk into their shop is a big enough player in the market or if it has a lot of success within this business, this could potentially inspire other businesses to use this technology as well. It would also provide a source of reference to other interested parties, since they could then also see, that the application works flawlessly.

7.4.2 C2C

In a C2C environment one thing is vastly different. The absence of businesses. So synergistic effects are likely not going to matter in this transaction context. Instead one of the most powerful ways of promoting the application, is good old-fashioned advertisement. More specifically the word of mouth type of advertisement. But while this kind of advertisement is very cost efficient, it does require some amount of people, that already know about the product, to really take off. But if it is possible to create such an initial core of satisfied users, they might grow passionate of the product and advertise the application better than any conventional campaign could. So there is a considerable incentive to put in this initial effort.
How does one go about creating such a strong core of users? Targeting the right people is definitely a good starting point. So who would be the "right" people in this context? As mentioned in the previous section, users with privacy and/or security concerns could very well constitute a desirable selection of individuals. How to find these people is an entirely different affair. Possible methods to achieve this could include: deploying adverts on relevant websites (i.e. that cater to this kind of audience) or hanging posters in buildings that are occupied by computer science departments of universities.
Ways to promote the application within a more general population likely require greater efforts. For most people, using CoinBlesk or a similar application, instead of more a conventional products has no real benefit for them. So they could to be incentivized through similar means as described in the B2C section, since they address the same people. Specifically for the C2C transaction context another way to promote the product could be using principles of gamification [20]. For example implementing some sort of achievement system [21] known from popular video games. Such a system would probably need to be implemented locally on the clients, so that privacy can still be guaranteed.

7.5 Summary and Conclusion

In this seminar thesis, the Bitcoin technology, CoinBlesk, as well as different mobile payment solutions were discussed. Additionally, the problems, that are currently hindering this application from becoming more popular, were also a subject of discussion. A few

possible solutions have also been presented. Unfortunately there seems to be no ideal solution and a solution depends heavily on the goals of the developers of this program. If the goal is to establish the application as a niche product for a specific group of users, the ”best” route to take differs greatly from the one where the goal is to make the product as popular as possible. The routes to take from here on out also depend on the resources, both financial as well as time related, the developers are willing to allocate for this project. The authors of this text believe that the appropriate goal should be to establish the application within a niche group of users. This course of action would allow the project to grow organically, because the system has a lesser chance of being suddenly overloaded. Also adjustments to the system could be made in a more effective manner, since this route would also provide the developers with more useful feedback. The step from an established product with a strong user base to mass product would certainly be more feasible than if one would try to aim for a mass product directly.

Bibliography

[1] Crypto-Currency Market Capitalizations; http://coinmarketcap.com, retrieved 10.2015

[2] Bitcoin is an innovative payment network and a new kind of money; https://bitcoin.org/en/, retrieved 10.2015

[3] Satoshi Nakamoto: Bitcoin: A Peer-to-Peer Electronic Cash System, 2009

[4] Joshua Davis: The Crypto-Currency, October, 2011; http://www.newyorker.com/magazine/2011/10/10/the-crypto-currency, retrieved 10.2015

[5] Stefan Mey: Ripple, Litecoin und Co.: Die wichtigsten und wertvollsten Bitcoin-Alternativen, February, 2014; http://www.spiegel.de/netzwelt/web/alternativen-zu-bitcoin-kryptowaehrungen-boomen-a-955763.html, retrieved 10.2015

[6] Bitcoin Developer Guide; https://bitcoin.org/en/developer-guide#mining, retrieved 10.2015

[7] Haeufig gestellte Fragen - Wie werden bitcoins erzeugt?; https://bitcoin.org/de/faq#wie-werden-bitcoins-erzeugt, retrieved 10.2015

[8] Bitcoin open source implementation of P2P currency; http://p2pfoundation.ning.com/forum/topics/bitcoin-open-source, retrieved 10.2015

[9] Raphael Voellmy: Averting the tragedy of the Blockchain, January, 2013; http://www.academia.edu/8865335/Averting_the_tragedy_of_the_Blockchain, retrieved 10.2015

[10] Raphael Voellmy: CoinBlesk, a Mobile NFC Bitcoin Payment System; Bachelor Thesis, August, 2015, Zurich

[11] Development of the bitcoion price in euro; https://www.bitcoin.de/en/chart, retrieved 11.2015

[12] MyWallet - Ihr digitales Portemonnaie; https://www.t-mobile.de/apps-und-partner/mywallet/0,28006,28533-_,00.html, retrieved 11.2015

[13] Number of MyWallet users; https://blockchain.info/charts/my-wallet-n-users?timespan=all&showDataPoints=false&daysAverageString=1&show_header=true&scale=0&address=, retrieved 11.2015

[14] Number of transactions per day; https://blockchain.info/en/charts/n-transactions?timespan=all&showDataPoints=false&daysAverageString=1&show_header=true&scale=0&address=, retrieved 11.2015

[15] Joseph Bonneau, Andrew Miller, Jeremy Clark, Arvind Narayanan, Joshua A. Kroll, Edward W. Felten: *SoK: Research Perspectives and Challenges for Bitcoin and Cryptocurrencies*, 36th IEEE Symposium on Security and Privacy (SP 2015), San Jose, USA, May, 2015 http://www.ieee-security.org/TC/SP2015/papers-archived/6949a104.pdf

[16] Reto Stauffacher: Selfscanning "Hab's probiert, aber mir fehlt eine Hand", October, 2015 ;http://www.beobachter.ch/konsum/kauf-leasing/artikel/selfscanning_habs-probiert-aber-mir-fehlt-eine-hand/, retrieved 11.2015

[17] Jonathan Todd Barker: Article on Investopedia.com, May, 2014; http://www.investopedia.com/articles/investing/052014/why-bitcoins-value-so-volatile.asp, retrieved 11.2015

[18] Jonas Fehling: Deutsche fÃ¼rchten Hacker-Angriffe auf Bitcoins;http://tiny.uzh.ch/vq, retrieved 11.2015

[19] Mark B. McKinley: *Psychology of Collecting*, http://nationalpsychologist.com/2007/01/the-psychology-of-collecting/10904.html, retrieved 11.2015

[20] Huotari, K., & Hamari, J.: *Defining Gamification - A Service Marketing Perspective*. Proceedings of the 16th International Academic MindTrek Conference 2012, Tampere, Finland, October, 2012

[21] Hamari, J., & Eranti, V. (2011). Framework for Designing and Evaluating Game Achievements. Proceedings of Digra 2011 Conference: Think Design Play, Hilversum, Netherlands, September, 2011

[22] Paul Gerhardt Schierz, Oliver Schilke, Bernd W. Wirtz *Understanding consumer acceptance of mobile payment services: An empirical analysis*. Electronic Commerce Research and Applications, 2009, 209-216

[23] Kalina S. Staykova, Jan Damsgaard *The race to dominate the mobile payments platform: Entry and expansion strategies*. Electronic Commerce Research and Applications, 2015, 319-330

[24] CoinBlesk https://bitcoin.csg.uzh.ch/, retrieved 11.2015

[25] Tapit https://www.tapit.ch/de, retrieved 11.2015

[26] Tapit - Pressemitteilung https://www.tapit.ch/sites/default/files/downloads/files/tapit_pressemitteilung.pdf, 01.07.2014, retrieved 11.2015

[27] Blick - Tapit wird 2016 eingestellt. http://www.blick.ch/news/wirtschaft/tapit-wird-2016-eingestellt-aus-fuer-swisscoms-bezahl-app-id4042585.html, 05.08.2015, retrieved 11.2015

[28] Mobile Pay - Danske Bank http://www.mobilepay.dk/, retrieved 11.2015

[29] Mobile Pay - The story in English http://www.mobilepay.dk/da-dk/Pages/The-story-in-English.aspx, retrieved 11.2015

[30] Mathias Ohanian: Bargeldlose Schweiz: Vorbild Daenemark zeigt, wie es geht http://www.handelszeitung.ch/lifestyle/digital/bargeldlose-schweiz-vorbild-daenemark-zeigt-wie-es-geht-780902, May, 2015, retrieved 11.2015

[31] Paymit - UBS https://microsites.ubs.com/paymit/de/, retrieved 11.2015

[32] Paymit - Medienmitteilung: SIX, UBS und ZÃ¼rcher Kantonalbank lancieren Bezahl-App Paymit, April, 2015 https://microsites.ubs.com/paymit/de/medien.html?file=files/UPM/documents/0428-d-p2p.pdf, retrieved 11.2015

[33] Start von Bezahl-App fuers Handy gut angelaufen http://www.tagesanzeiger.ch/wirtschaft/geld/Start-von-BezahlApp-fuers-Handy-gut-angelaufen/story/11499959, June, 2015, retrieved 11.2015

[34] Katherine Fletcher: Denmark headed toward cashless society https://coinreport.net/denmark-headed-toward-cashless-society/, June, 2015, retrieved 11.2015

[35] Ralf Beyeler: Tapit: App macht Smartphone zur Kreditkarte. https://www.comparis.ch/telecom/mobile/news/2014/07/tapit-swisscom-app.aspx, July, 2014, retrieved 11.2015

[36] Beat Bernet, *Aspekte der Technologieakzeptanz von Mobile Payment Services.* http://link.springer.com/chapter/10.1007/978-3-642-54411-8_14, Wirtschaftsinformatik in Wissenschaft und Praxis, Business Engineering, Springer Berlin Heidelberg, 2014, 193-203

[37] Denis Dennehy, David Sammon *Trends in mobile payments research: A literature review.* Journal of Innovation Management, 2015. JIM 3, 49-61

Chapter 8

Certificates

Noah Zani, Tim Strasser, Andrés Baumeler

Certificates are widely used in digital communication to ensure the identity of a communication partner and to prove the ownership of a public key. Certificates are an important part of the Public Key Infrastructure, which is used to provide privacy and data integrity between two communicating computer systems. The rising amount of sensible data like payment information sent over the internet has caused discussions about the role, importance and legitimacy of communication security, data surveillance and cryptography in general in the new digital age, also contributing to the rising economic importance of certificates and their providers in the last few years. This report will lay out how Certificates are issued, revoked and distributed by Certification Authorities. Further this report will introduce current and future challenges faced by the various Authorities that are part of the Public Key Infrastructure and present some approaches to solve these challenges.

Contents

8.1 Introduction **171**
8.2 Fundamentals of Cryptography **171**
8.2.1 Symmetric Cryptography 171
8.2.2 Asymmetric Cryptography 172
8.3 Public Key Infrastructure **173**
8.3.1 Certificate Authority 174
8.3.2 Hierarchical Certification 175
8.3.3 Certificate Revocation 177
8.3.4 Web of Trust 179
8.4 X.509 Overview **179**
8.4.1 X.509 Certificate Structure 180
8.5 RFC 5280/6818 Overview **182**
8.5.1 Comparison of certificate extensions 182
8.6 Transport Layer Security **183**
8.6.1 TLS Handshake 184
8.6.2 Cipher Suites 187
8.7 Economic Aspects of Certification **187**
8.7.1 Commercial Certificate Authorities 187
8.7.2 Free Certificate Authorities 188
8.7.3 Private Certificate Authorities 188
8.7.4 Choosing a Certificate Authority Type 189
8.8 Summary **189**

8.1 Introduction

This report will first introduce you to the basics of symmetric and asymmetric cryptography and show how those concepts are used by certificates. The second part will introduce the public key infrastructure - the infrastructure needed to issue, revoke and distribute certificates in a reliable way - and its major components. The last part of this report will discuss some economical aspects of public key infrastructures, especially with regard to certificate authorities, the entities in charge of issuing certificates.

8.2 Fundamentals of Cryptography

This section will give a short introduction to the two main types of cryptography, symmetric cryptography and asymmetric cryptography and cover some advantages and disadvantages that come with each type of encryption and, using these points as a basis, introduce the need for certificates and the benefits they bring.

8.2.1 Symmetric Cryptography

Symmetric encryption respectively symmetric key algorithms are a concept of cryptography in which the same cryptographic key is used to encrypt and decrypt the plain text of the message to be communicated.

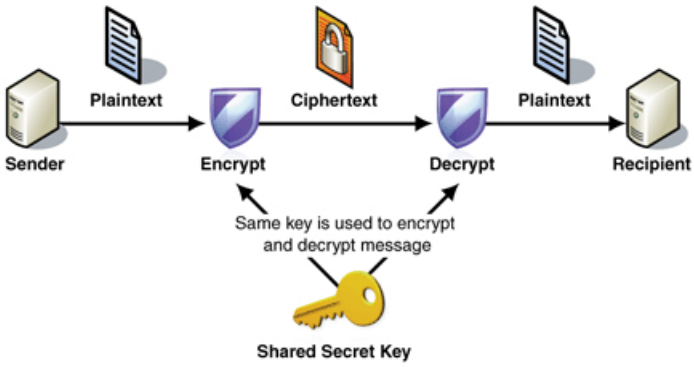


Figure 8.1: Communication between two parties using symmetric encryption i.e. a shared secret key [20]

While this form of encryption is fast and simple due to a single shared secret key being used, one great disadvantage is that all parties participating in the communication have to have knowledge about the shared secret key. Because the key is needed to secure the communication channel, it cannot be sent over the unsecured channel. If there already were a channel on which the parties could exchange the encryption key securely, encryption would not be needed in the first place. This problem is called the key exchange problem. Historically, the parties exchanged parts of key via a secure channel by using a trusted courier. While the key exchange was tedious and costly, the rest of the communication could be performed over insecure channels like the postal service. Another drawback is that if the single shared secret key is compromised, the plain text of every piece of information shared between the communicating parties can be read. The same key is only used once, every other use makes the communication insecure. The encryption of messages can be either done by stream ciphers or block ciphers. A cipher being an algorithm used for encryption or decryption. Using stream ciphers, the digits (typically the bits) of the message plain text are encrypted one at a time by performing an operation on the plain text digit and the corresponding digit of the key stream, which is generated by another algorithm using the shared key as the input (seed value). Block ciphers perform an

algorithm on a block with a predefined bit size of the plain text, taking the block and a secret key as the input and returning a cipher text block of the same bit size. One of the most important symmetric encryption algorithms, in part because of how widely used it is, is AES (Advanced Encryption Standard) a.k.a. Rijandel, a 128-bit block cipher which was developed by order of the U.S government as a successor of DES (Data Encryption Standard). Worth mentioning is also IDEA (International Data Encryption Algorithm), a block cipher developed by ETH Zurich that is used by the PGP (Pretty Good Privacy) program and the Web of Trust developed by Phil Zimmerman. For more details on PGP and Web of Trust, see Subsection 8.3.4. Some other well known stream ciphers are SEAL (Software-Optimized Encryption Algorithm), a very fast stream cipher well suited for encrypting hard drives and the even faster Quick Stream Cipher (QUISCI).

8.2.2 Asymmetric Cryptography

In asymmetric encryption, the parties exchanging messages all hold a pair of keys consisting of a private and a public key. The sender encrypts his message with the recipients public key. The recipient can decrypt the data using his corresponding private key.

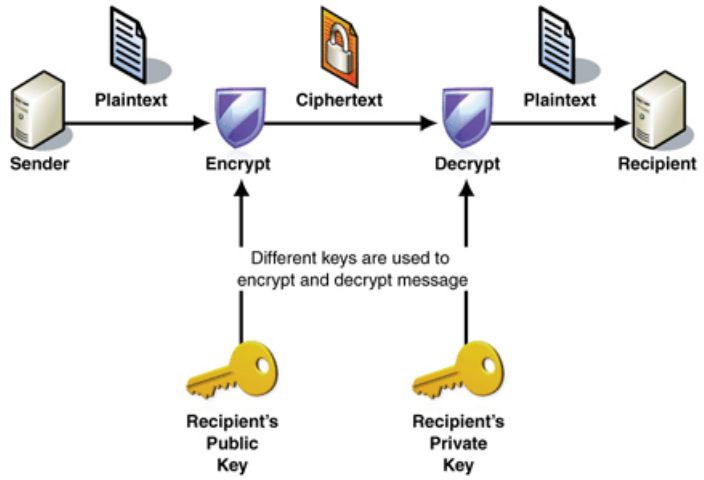


Figure 8.2: Communication between two parties using asymmetric encryption [20]

The advantage over symmetric encryption is that with this concept, no previous communication between the sender and the recipient is required to exchange a key for encrypting the communication flow. The key exchange problem is therefore solved by making the encrypting key public and eliminating the need to exchange it securely. Also, a compromised private key does not leave all messages vulnerable, only data sent to the party whose private key is known can be decrypted by other parties.
Using this type of encryption, the only main problem left is that the communication is based on the parties trusting in each others identity i.e. relying on the fact that the receiving party is indeed the entity it claims to be and that the public key belongs to that party. This problem of trust is solved by certificates binding a public key to an entity. In essence, a certificate is a file confirming the ownership of a public key by an entity and a signature by a third party on whose authority to issue certificates both communication partners agree.
It becomes apparent that a certain infrastructure is needed for the management of issuing certificates and related services. This infrastructure is called a public key infrastructure (PKI). The following section will further discuss the concept of a PKI, its elements and functionality.
In practice, the public key of the parties are not used directly for to secure communication i.e. for encrypting and decrypting messages. Instead public key cryptography in

combination with certificates is used to securely communicate a shared session key which is then used to encrypt further communication in a symmetrical manner. This approach is discussed in detail in Section 8.6 explaining how the TSL protocol is used to establish a secured connection between a client and a web server. The session key is only valid for the current session and is not used for future sessions between the same two parties. The validity of the public and the private key is confirmed by the certificate, which typically is valid for one and a half years to two years, but it can be valid longer. Validity periods are not a consequence of technical restrictions, they can be chosen arbitrarily by the issuer of a certificate.

8.3 Public Key Infrastructure

Certificates are needed for reliable and secure asymmetric encryption. But certificates are only worth as much as they can be trusted the issuer of the certificate. To ensure trust, two things are necessary. On one hand, you have to trust the entity that issued the certificate and on the other hand, you have to ensure that the certificate actually is from said entity. Trusting a certificate issued by someone you do not know or is not verified is as good as trusting someone without a certificate at all. Because of this the public key infrastructure (PKI) is built around asymmetric encryption. Digital Certificates are just one part of the whole infrastructure. A PKI is defined as:

> ”The infrastructure able to support the management of public keys able to support authentication, encryption, integrity or non-repudiation services.” [15]

A PKI consists of several distinct entities, namely the Certification Authority (CA), the Registration Authority (RA), and the Validation Authority (VA). Those are logically independent entities as they fulfill different purposes. Though they do not need to be separate in practice. In fact, it is quite common for a CA to also function as a RA and even VA.
Figure 8.3 shows the processes needed to issue a Certificate. If user A wants to securely communicate with user B by relying on a PKI, he needs a digital certificate that includes his public key and information and states his identity. To get this certificate he needs to send in an application for a certificate to the RA. The RA then verifies the requester and the data supplied. If the data meets the requirements of the RA, the request is forwarded to the CA, which then creates a certificate with the given data and renders it tamper-proof by signing it with its own private key. This certificate is then issued to user A, who in turn sends it to user B so user B can verify the identity of user A. Before accepting the certificate, user B first checks its validity. This is done by sending a request to the VA, which can tell if the certificate is valid, because it has this information from the CA. When user B gets a positive response, a secure communication can be established. If not, user B would reject to communicate with user A. User A could then start a new request, but with the same certificate the result would not change.

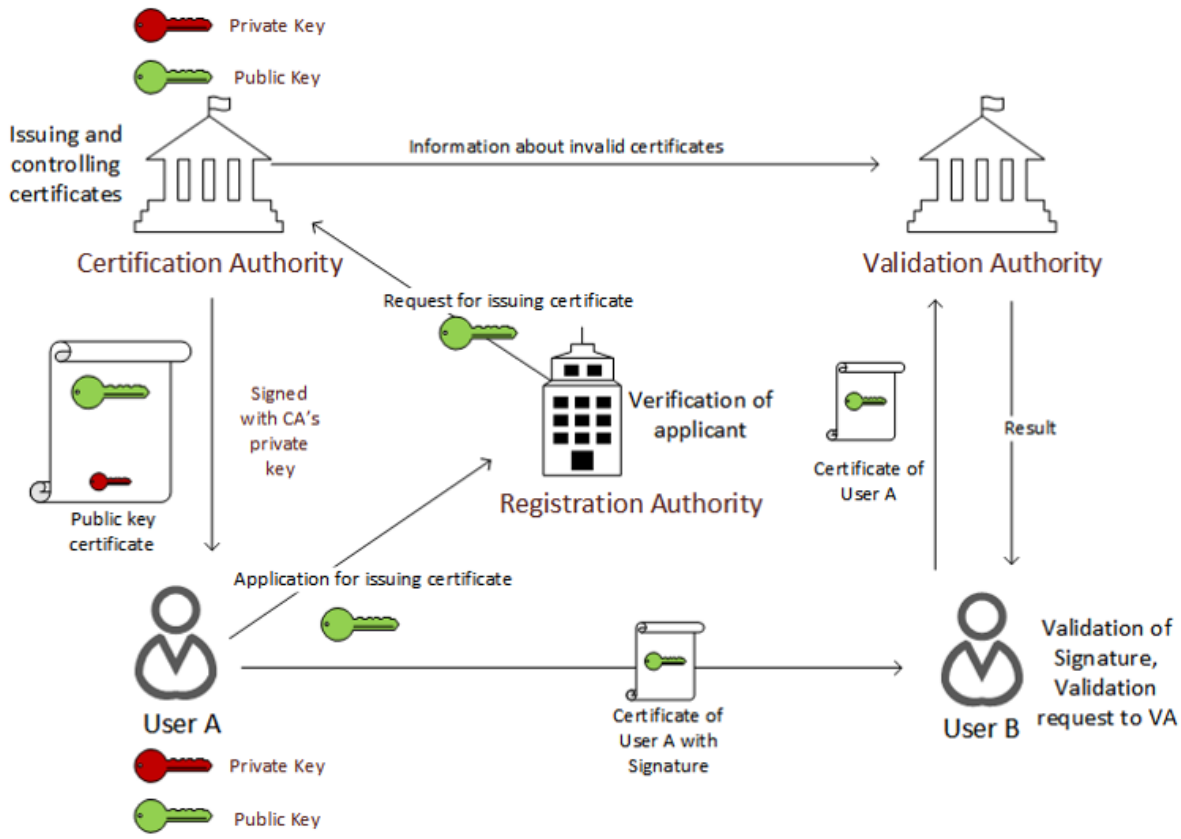


Figure 8.3: The entities of a PKI and how they work together

8.3.1 Certificate Authority

Definition:

> "An authority trusted by one or more users to create and assign public-key certificates. Optionally the certification authority may create the subjects' keys." [15]

A Certification Authority (CA) is the entity of a PKI that actually issues the digital certificates. It takes the role as a trusted third-party. That means, if for example a user wants to access and share critical data with a website, he relies on a certificate of a trusted CA. In its core, the issued certificate contains information about the certificate holder (e.g., name, and his public key). The certificate is signed by the issuing CA with its own private key to render it tamper-proof. This signature can then be checked with the public key of the issuing CA. Due to this role as trusted third-party, it is crucial that CAs are absolutely trustworthy, because otherwise it would not generate any benefit. To ensure this trust they invest a lot in security to protect their private key and systems overall by implementing various IT security measures. They also undergo annual security audits often by multiple institutes like ETSI or WebTrust. Those audits are done by certified companies, often consulting or advisory companies, that then visit the CA and check numerous criteria defined by the according standard (e.g., WebTrust).
Almost all of modern secure communication relies on CAs, but they are not trustworthy beyond doubt. They are arguably the most critical point in terms of security threat. Because once a CA would be subverted, potentially the whole system and all the entities that trusted this particular CA could be subverted as well. In fact there are several known cases where a CA got compromised. One mayor scandal was in 2011, when the Dutch company DigiNotar was compromised and the attacker issued several hundred fraudulent certificates [17]. This became public, because a wild card certificate for Google was used for attempted man-in-the-middle attacks. The attacker tried to get between Google

users mainly from Iran and encrypted Google services [18]. In response most browser providers blacklisted all DigiNotar issued certificates. The Dutch government took over operational management and an investigation was started, but the company was then declared bankrupt the same month. This underlines once more, that trustworthiness is absolutely key for a CA. But even if more compromised certificates would have been out of question and trust would have been rebuilt, it is a highly controversial question, if governments should be involved or run CAs. Arguments against it is the trustworthiness of governments themselves (corruption, abuse of power e.g. the controversy around the National Security Agency (NSA) of the United States) or that they would not necessarily have the economic incentives to stay trustworthy and be always up to date with the latest technology.

Prior to issuing a certificate, the CA has to assure it issues them to the correct user, computer or other entity. This step can be delegated to a Registration Authority.

The RA is an optional instance in a PKI is working together closely with the CA. [1] Its main task is to receive a certificate signing request (CSR), validate the authenticity of the user or organization requesting the certificate and check the correctness of the information in the CSR, and then forward the request to the certification authority that issues the final certificate. Authentication at the RA can be done by software or manually, typically involving checking personal papers under human supervision.

The thoroughness of this check depends for one on the purpose of the certificate that is requested. As example for a SSL certificate used for https connections, it is generally distinguished between Domain Validation, Organizational Validation and Extended Validation.

Another responsibility of the CA vital to maintain its trustworthiness is to revoke its issued certificates if they are no longer valid. Reasons for revoking among others can be a compromised key and end of specified lifetime. A CA can publish those revoked certificates itself or mandate a Validation Authority with this task.

The validation authority provides this service, either through providing a certification revocation list (CRL) as download or an interface that enables real-time checking through protocols like the Online Certificate Status Protocol (OCSP) or the Server-based certificate validation protocol (SCVP). The concept of certificate revocation and status protocols will be explained in the upcoming section.

8.3.2 Hierarchical Certification

Typically, certificates are used in a strictly hierarchical public key infrastructure. The trust model of such a PKI is a tree-like structure and requires a highly secured entity at the top of the chain called a root certification authority, which is trusted by all the participants and represents the trust anchor of the PKI. The certificate that a root CA holds is called the root certificate and is self-signed i.e. issued by and to the root CA itself. The root certificate is used to issue certificates to certification authorities directly subordinate to the root CA called intermediate authorities. They inherit the trustworthiness of the root certification authority. Any certification authorities lower in the hierarchy than the root CA are called subordinate CAs. For example, the university of Zurich offers a service together with SWITCH that enables students to get a certificate for encrypting and digitally signing emails. The university acts as a RA, checking personal papers of the person applying for a certificate and confirming the identity. After checking, the request is forwarded to SWITCH. SWITCH offers the service of a CA and issues the certificate. The CA of SWITCH itself is certified by a CA of Quovadis, a Certification Service Provider. The CA of SWITCH is subordinate to the CA of Quovadis, which itself is subordinate to the root CA in the CA hierarchy of Quovadis. He hierarchical structure might reflect organizational structure with intermediate CAs being responsible

for departments or countries in infrastructures of a larger scale. On the lowest level of the hierarchical structure are the issuing certificate authorities, which issue certificates to end entities such as users, devices, servers and services.

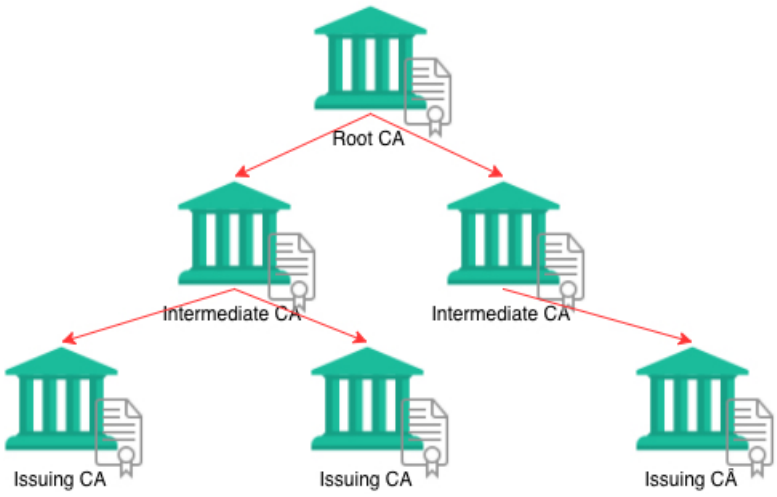


Figure 8.4: Hierarchic relationship between certification authorities

Certification Path The hierarchical structure is reflected in a certificate chain or certification path and is essentially a series of certificates issued by the successive certification authorities in the hierarchy. The most common and thus important root certificates are preinstalled in browsers and operation systems in a root store. Certificates can also be manually installed on entities in a public key infrastructure. This is commonly done in companies and other organizational PKIs. For example, Figure 8.5 shows a screenshot of the keychain access app in OSX El Capitan. The application lists all certificates that are installed on the device, their current validity and their expiration date. The list shows that root certificates of two of the major companies that issue certificates, VeriSign and Thawte, are installed on the system.
In order to establish a connection with an entity presenting a certificate that is not explicitly trusted, the certification path has to be validated starting from a trust anchor. The trust anchor is the entity in the certification path that is trusted explicitly beforehand and from which the chain of trust is derived. RFC5280 - a specification by the Internet Engineering Taskforce (IETF) - specifies a certificate path validation algorithm that takes the path to be validated and other parameters such as the current time and the trust anchor of the path as an input and then traverses the path, starting from the trust anchor and checking the validity of the chain through several steps, including comparing the current time against the validity period of the certificates, checking their revocation status using a certificate revocation list or a service like OCSP, both explained in Section 8.3.3. The algorithm verifies that given a path of n certificates, the following conditions hold: [1]

```
(a)  for all x in {1, ..., n-1}, the subject of certificate x is
     the issuer of certificate x+1;
(b)  certificate 1 is issued by the trust anchor;
(c)  certificate n is the certificate to be validated (i.e., the
     target certificate); and
(d)  for all x in {1, ..., n}, the certificate was valid at the
     time in question.
```

Listing 8.6: Conditions for path validation as desribed in RFC5280 [1].

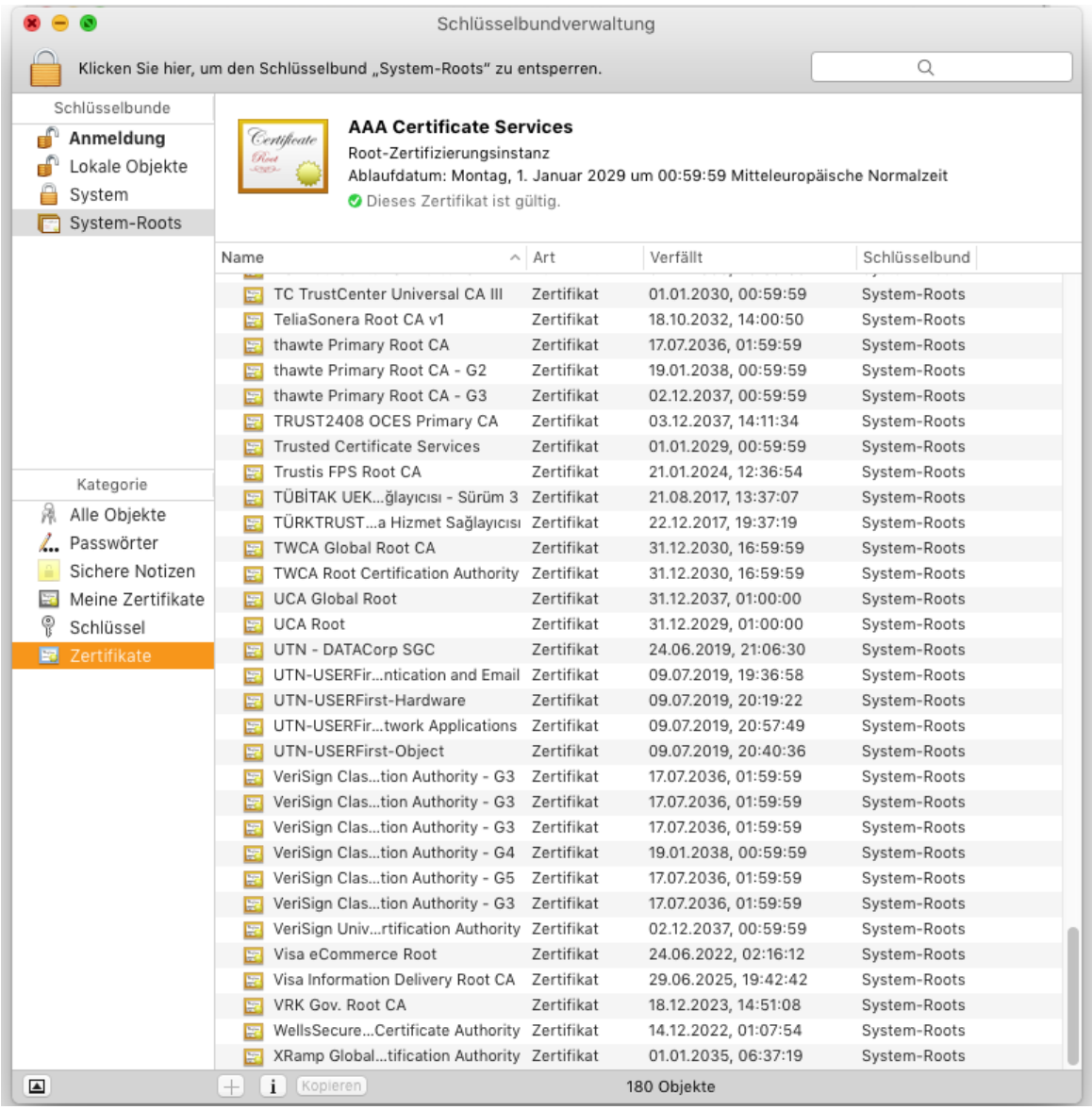


Figure 8.5: Screenshot of the store of trusted root certificates in OSX El Capitan

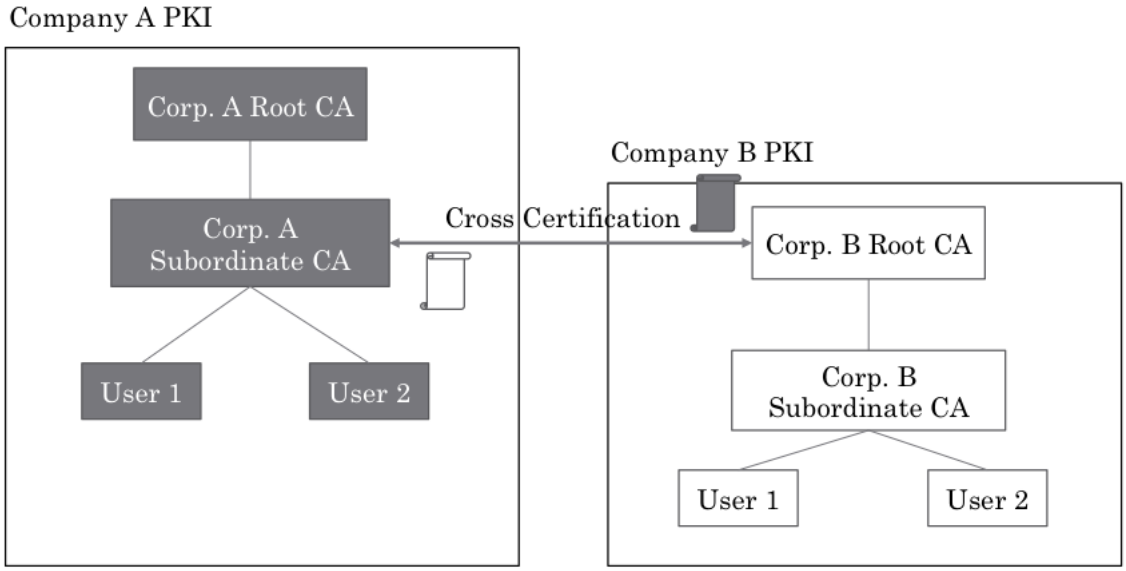
The certification path validation algorithm ends successfully if no error condition is found until the last certificate in the chain is reached. In case of a failure, the process terminates with output giving a reason for the failure [22] [1].

Cross-Certification

Cross certification is a concept that enables entities of two independent PKIs to collaborate in a secured manner. For the connection to be secure and trusted, two CAs, each part of a separate PKI issue a cross certificate, confirming that they trust each other and perform a mutual agreement about the exact policies used in communication between CAs. In contrast to the hierarchical structure in a single PKI, the two CAs participating in a cross certification are on the same level i.e. there is no subordinate CA in the trust relationship, even if the two CAs engaging in the relationship are on different levels in their respective PKI, see Figure 8.7.

8.3.3 Certificate Revocation

If a certificate is found to be invalid, which can be the case if it is discovered that a certificate was not properly issued or if there exists a suspicion that the private key has been compromised, it is called "revoked". While the status of revocation is irreversible, a certificate status can also be set to a state of hold, making the certificate only temporarily



Listing 8.7: Cross Certification between two independent organizational public key infrastructures

invalid. This might happen if a user is not sure if his private key has been stolen or not. In the X.509 standard, revocation of certificates are managed by the CA publishing a signed and time-stamped certificate revocation list (CRL), respectively enabling the download of the list through a public repository. It is also possible for the CA to delegate the task of publication to another trusted authority. Since a CRL is signed and, thus, its origin can be validated the distribution can be carried out via an unsecured communication [1]. Publication can happen periodically at a defined interval or after a certificate has been revoked. The validity of the information in a CRL is commonly given during a time frame of 24 hours. There are several disadvantages coming with the use of CRLs. One is the potential large size of a list, which led to the idea of incremental lists also called Delta CRLs, only representing the changes to the CRL last published. Another problem is that even if publishing intervals are kept short, CRLs are unable to provide real-time information about a certificate's validity, therefore there is always the risk of a certificate being revoked and not yet being mentioned in a CRL, leaving communication vulnerable for up to several hours. Also, failing to retrieve the current CRL might lead to a certificate falsely believed to be valid.

An alternative to CRLs is the use of a protocol like OCSP, enabling real-time checking of a certificate's status. OCSP was first introduced in 1999 in RFC2560, its current version is described in RFC6960[13][14]. Using OCSP, it is possible for a client to obtain information about a single certificate's validity by contacting a server, commonly called a OCSP responder and run by the certificate issuer, with a request containing among other fields the protocol version and the target certificate identifier.[14] After receiving a request and checking the request for conforming to the standard and containment of all information needed, the OCSP responder sends back a signed response containing sub-responses for every certificates requested, stating again the target certificate identifier, the status value and the validity interval during which the client can rely on the information received from the responder. The status value of a certificate can be either "good", "revoked" - which includes certificates put on a hold status - or "unknown" [14]. While this type of certificate status checking removes a big burden off from the client's network, it can put stress on the CAs side, for example for a website that is visited in high frequency from clients all requesting information about the certificate that the CA has issued for that website. The need of contacting a third party during the communication is also a privacy concern, the CA now has knowledge about the time of the visit and the website visited. This problem has been solved through a concept called "OCSP stapling" used

in SSL/TLS handshakes, where the issue holder regularly contacts the CA in order to retrieve a timestamped validation of its certificate status which can be shown to clients. This removes the need for the client to contact the CA directly.

8.3.4 Web of Trust

The web of trust is a decentralized alternative trust model to the centralized, hierarchical trust model of a PKI. It was first introduced by the creator Phil Zimmermann of Pretty Good Privacy (PGP), a program used for encryption, decryption and signing of data communication, often used for e-mail encryption [10] [12]. In contrast to a centralized approach, which involves a certification authority acting as a trusted third party, a web of trust is based on mutual trust amongst its nodes. In the PGP user manual Zimmermann writes:

> "As time goes on, you will accumulate keys from other people that you may want to designate as trusted introducers. Everyone else will each choose their own trusted introducers. And everyone will gradually accumulate and distribute with their key a collection of certifying signatures from other people, with the expectation that anyone receiving it will trust at least one or two of the signatures. This will cause the emergence of a decentralized fault-tolerant web of confidence for all public keys." [11]

This means that each person has a collection of signatures from other people directly trusting that person and can express their trust in another person through signing that person's public key. This creates two types of trust, totally independent from each other: First, trust in the fact that a person is who he claims to be (authenticity) and second, trust in the fact that the other person is careful while signing other people's public keys. Signing other people's public key is done at so called key signing parties. A key signing party is an event common in the PGP community where participants mutually validate each others authenticity on site. This is done by checking identifying documents like a passport. Before the event, each participant generates a hash function of his public key. The output of the hash function is short enough to be copied and written down by other participants. After the event, the participants reconstruct the whole public keys from the shortened keys that they have written down using the hash function that was originally used to shorten the keys and digitally sign them.

8.4 X.509 Overview

In this section one of the most important, since most widely distributed, recommendations in the field of digital certification is explained. The X.509 Recommendation, published by the The Telecommunication Standardization Sector of the International Telecommunication Union (ITU-T) was developed as part of the X.500 standard, which described the structure of a globally accessible directory service but was never fully implemented. Nevertheless it is still of high relevance as of today, because it has a life outside of the remaining X.500.
It was first published on the 25th November 1988 by the ITU-T. Over the years several updated editions got published, the latest being edition 7.0 from the 14th October 2012. [15]
Version 1 defined a fixed set of fields for a certificate, like the issuer which contains the name of the issuer, the validity which specifies the lifetime of the certificate or the serialNumber which identifies the certificate, among others. Further it assumed a hierarchical system of certification authorities. Version 2 was an enhanced one that added some more

fields. The 3rd version then came with more improvements. For one it allowed for alternative topologies like meshes and bridges. And it also allowed for certificate extensions. Those consist of a type-and-value pair, with an optional critical flag. One often seen use for extensions is to issue a certificate only for a specific purpose.
The X.509 recommendation addresses multiple security requirements in areas like authentication and other security services. It accomplishes that by defining frameworks for public-key certificates and attribute certificates.

> "The public-key certificate framework is the base specification for public-key certificates, for the different components going into a public-key infrastructure (PKI) for validation procedures and for public-key certificate revocation, etc. The attribute certificate framework is the base specification for attribute certificates and the different components going into the Privilege Management Infrastructure (PMI)." [15]

In the end, X.509 is about message integrity, authentication and authorization. Integrity, because it tries to ensure that a certificate can not be compromised as long as the signing private key is secure, since this can be checked by the receiver of a certificate. Authentication, because the CAs must ensure, that the certificate holder is the one stated in the certificate. Authorization due to the applications in a Privilege Management Infrastructure, which is not part of this paper.
These frameworks can be adopted by whoever wishes to. All of these recommendations are voluntary and can be followed or not. However, there are certain mandatory parts that must be followed to be in compliance with the Recommendation as such. One example of a mandatory requirement is, that the before mentioned serialNumber field must hold an integer assigned by the issuing CA and that it must be a unique value for this CA.
It is the basis or gets at least supported for many other specifications, such as the secure socket layer SSL and its successor the transport layer security TLS protocol or the IETF PKIX. Others are IPSec, S/MIME, Smartcard, SSH, HTTPS, LDAPv3 and EAP.

8.4.1 X.509 Certificate Structure

This section is intended to give you an idea how a digital certificate looks like based on the ITU-T recommendation X.509 version 3 as specified in the IETF profile RFC5280. The X.509 recommendation is explained in detail in Section 8.4. Because the ITU-T X.509 standard is very broad, the RFC3280 Profile was developed to adapt the X.509 system for use in the internet. The RFC3280 was updated by the RFC5280 Profile in 2008 [1].
The basic fields of a X.509 v3 certificate consists of the version field, serial number field, issuer field, validity field, the subject field the subjects public key and the signature. The version field is used to identify with which X.509 version a certificate complies. The serial number field contains a positive integer value assigned by the issuing certificate authority. It must be unique for each certificate issued by a single certificate authority. The signature algorithm field contains the algorithm used by the certificate authority to sign this certificate. The issuer field identifies the entity who has signed and issued the certificate. The validity field specifies the certificates validity period. The validity period is defined as the time the issuing CA will maintain information about the status of the certificate. The subject field identifies the entity associated with the public key stored in the public key field. The issuer and the subject field must follow the naming specifications defined in RFC3280. The subject public key info field contains the public key of the subject and the algorithm it is meant to be used with. After the subjects public key field follows the certificates signature. The signature is preceded by an signature algorithm identifier field. This field must contain the same algorithm as specified in

```
x509 -in fugaCert.pem -noout -text
Certificate:
    Data:
        Version: 3 (0x2)
        Serial Number: 1 (0x1)
        Signature Algorithm: sha1WithRSAEncryption
        Issuer: C=se, ST=skane, L=lund, O=lund university, CN=root ca
        Validity
            Not Before: Oct 15 15:26:47 2008 GMT
            Not After : Oct 15 15:26:47 2010 GMT
        Subject: C=se, ST=skane, O=lund university, OU=eit, CN=fuga
        Subject Public Key Info:
            Public Key Algorithm: rsaEncryption
            RSA Public Key: (1024 bit)
                Modulus (1024 bit):
                    00:9c:44:f3:46:5a:bf:fb:59:fe:e8:78:fd:da:73:
                    e4:96:ba:38:68:50:c5:fc:45:da:5b:12:5e:36:22:
                    93:a4:d8:1f:4b:83:65:da:48:f9:0e:52:13:46:25:
                    53:d1:f9:c0:15:99:68:2a:1f:2f:3d:9c:32:6c:4b:
                    9c:58:44:c3:50:ab:3e:b2:f6:a0:10:53:f3:86:f6:
                    7b:c2:53:0e:1a:2c:57:fb:12:d0:b6:da:53:df:d2:
                    34:cc:80:79:82:ad:09:bf:19:70:48:20:69:59:e9:
                    82:66:71:4b:86:55:49:ef:64:e4:2c:db:34:1c:a9:
                    70:91:d9:c4:ff:dd:de:dc:cd
                Exponent: 65537 (0x10001)
        X509v3 extensions:
         X509v3 Basic Constraints:
           CA:FALSE
         Netscape Comment:
           OpenSSL Generated Certificate
         X509v3 Subject Key Identifier:

B4:ED:55:6D:BC:EF:B7:8F:D7:03:80:49:A0:D7:CA:FC:D1:8D:4E:84
         X509v3 Authority Key Identifier:

keyid:A3:EE:94:F9:05:9F:25:F4:F4:C1:E2:6F:F8:8D:5C:3A:7A:52:9B:28
    Signature Algorithm: sha1WithRSAEncryption
        90:3f:49:ce:c4:b4:0f:96:b5:62:b4:06:a4:03:fc:9b:9c:80:
        6d:c7:46:7d:ab:93:1c:6a:da:72:89:0e:08:7e:7d:44:f9:e1:
        f5:de:f3:96:ca:d4:6c:bb:9c:5a:4d:9b:cb:e0:e1:a7:f0:95:
        41:2d:f5:d3:90:92:33:cd:d9:41:64:a8:2f:38:fa:bb:08:e0:
        ec:e2:cc:db:0c:eb:f8:ad:8f:ba:52:3e:a3:20:21:ce:6a:a1:
        e9:93:5b:a6:70:b3:5c:0f:0d:63:c4:56:d8:fc:f7:fe:d3:5d:
        a6:7c:cb:d1:10:3d:da:70:eb:dd:70:a5:e5:c7:7c:1e:2f:4e:
        e7:5d:6f:2f:15:77:58:69:5a:4b:90:eb:86:6d:d5:00:06:a4:
        c9:02:79:f3:88:6f:7f:26:22:7e:03:2f:78:61:e9:5a:21:d9:
        f4:6c:6e:18:0b:9c:44:be:ee:9b:b3:d6:97:23:1a:32:8e:d3:
        66:e1:5e:c4:1d:d3:09:34:d5:7a:d0:af:88:a2:85:04:38:d1:
        c1:03:d5:55:af:90:d7:0c:fa:8d:0e:1d:0f:6e:95:b3:12:86:
        ee:c5:4c:31:c1:f1:47:59:7e:b7:a1:09:78:2f:7a:5f:1f:28:
        68:69:8b:d9:fd:b7:18:a1:35:bc:e7:95:0c:d2:a8:4a:19:5b:
        1f:c1:c6:51
```

Listing 8.8: Example of a Server Certificate [5]

the first signature algorithm field because it is used by a client to verify the certificates signature.

8.5 RFC 5280/6818 Overview

The purpose of the RFC 5280, the Internet X.509 Public Key Infrastructure Certificate and Certificate Revocation List (CRL) Profile, is as the name implies to profile the X.509v3 certificate and X.509v2 certificate revocation list for use in the Internet. That means, it takes these standards and describes and profiles them in detail including additional information regarding the format and semantics of certificates and certificate revocation lists (CRLs) for the internet. Also there is a description of possible certificate extension and a set of required certificate extensions is specified. A new possible extensions would be the one to provide means of identifying certificates that contain a particular public key (Subject Key Identifier Extension). Further there is an algorithm for X.509 certificatin path validation is described, which states that the certificate 1 has to be issued by a trust anchor, that for all x in 1, ..., n-1, the subject of the certificate x is the issuer of certificate x+1 etc. [1]
Goal of this specification RFC 5280 is, that the X.509 technology can be used in various applications such as surfing in the internet, e-mail, authentication and more. This is accomplished by the precise description and definition of algorithms, certificate and extension structures and more. Those can then be used in contrast to most of the contents of the X.509 recommendation, which can not be used as is because they often describe only the a concept rather than details. Depending on specific requirements, the RFC 5280 itself can be supplemented and extended again.

8.5.1 Comparison of certificate extensions

In recommendation ITU-T X.509, everything is quite vaguely and theoretical. It states that the extensions field allows for the addition of new fields to the structure without modification to the ASN.1 definition. But it does not limit what those fields could be. It only states that an extension field consists of an extension identifier, a criticality flag, and an encoding of a data value of an ASN.1 type associated with the identified extension. Going on it states that rules should be included, if sequence of extensions is of importance or if it does not matter. The section describes the differences of the critical/non-critical flag, which should be used, including how either one should be treated if it is not recognized by the receiver of the according certificate. A critical flag that is not recognized by the receiver should result in the rejection of the certificate. In contrast, when a non-critical flag is not recognized, this extension should be ignored, without resulting in the rejection of the whole certificate. Further the X.509 recommendation describes considerations regarding backwards compatibility, especially during a transition phase, when new extensions are used. It then states that any organization may define specific extensions depending on need and finishes by giving the following syntax to define specific extensions:

```
EXTENSION ::= CLASS {
          &id OBJECT IDENTIFIER UNIQUE,
          &ExtnType }
WITH SYNTAX {
          SYNTAX &ExtnType
          IDENTIFIED BY &id }
```

Listing 8.9: Extension as specified by X.509 [15].

The Listing 8.9 shows an information object class, that is used to define specific extensions. It states that a unique identifier is required as well as the type of the extension.
In RFC 5280 it is explained in detail how the extension works and how it looks like, at the example of the Key Usage extension. This extension, as the name suggests, defines for what purpose the key contained in in the certificate is intended. One use case could be, that the key should only be used for key management. Following the syntax of the extension with the possible restrictions:

```
id-ce-keyUsage OBJECT IDENTIFIER ::= { id-ce 15 }
KeyUsage ::= BIT STRING {
          digitalSignature (0),
          nonRepudiation (1), -- recent editions of X.509 have
                              -- renamed this bit to contentCommitment
          keyEncipherment (2),
          dataEncipherment (3),
          keyAgreement (4),
          keyCertSign (5),
          cRLSign (6),
          encipherOnly (7),
          decipherOnly (8) }
```

Listing 8.10: Extension as specified by RFC5280 [1].

In the first line of Listing 8.10, the identifier of this extension can be seen. In the second line is the value, which is a bit string called KeyUsage. This contains bits for the possible uses, for example a bit for digitalSignature on position 0, that can either be 0 (not asserted) or 1 (asserted). Depending on the requirement, the matching bits will be asserted. At least one of those bits must be set to 1.
Conforming CAs must include this extension in certificates that contain public keys that are used to validate signatures on other public key certificates or CRLs.
The bits as seen in the KeyUsage example above, are then described and what they affect as well as how they work together. An example: If the public key should be used to verify signatures on certificates as well as other objects, both the digitalSignature and the keyCertSign bits must be set. Some combinations, for example combining the nonRepudiation bit with other keyUsage bits, may also have security implications depending on the context in which the certificate is being used.
It was published in May 2008 by the Internet Engineering Task Force (IETF). This specification makes the RFC 3280 obsolete. The RFC 5280 got some updates itself in January 2013 by the RFC 6818.

8.6 Transport Layer Security

In this section the Transport Layer Protocol (TLS) is introduced and it is shown how TLS uses symmetric cryptography and mechanisms offered by the Public Key Infrastructure (PKI) to enable a secured connection between two communication partners. This section is intended to provide you with a real live use case of the PKI and Certificates introduced in the earlier sections. TLS is a protocol situated in the transport layer of the TCP/IP protocol stack. TLS is a proposed Internet Engineering Task Force (IETF) standard defined in RFC5246 [3] and last updated in RFC6176 [4] in the year 2011. TLS is based on the earlier Secure Sockets Layer (SSL) protocol specifications. SSL is deprecated and should not be used anymore because it does not provide a sufficiently high level of security and has known deficiencies.

8.6.1 TLS Handshake

The TLS protocol consists of two layers, the TLS Record Protocol and the TLS Handshake Protocol. This section will go over the TLS Handshake protocol, because this is the layer of the TLS Protocol that interacts with the PKI. The TLS Handshake protocol describes all steps needed to set up a TLS connection between two communicating computer applications (e.g. , a web browser and a web server). The setup of a TLS varies depending on how the client and the server are set up and on the version of the protocol in use. Because of this reason this section will introduce a more generalized version of a TLS handshake. For the remaining part of this section it is assumed that a client wants to establish as TLS secured connection with a web server. In this scenario only the server will be authenticated, but not the client. This is often the case when connecting to a website via https.
A basic TLS handshake can be divided into six steps [3]. In step one so called hello messages are exchanged to agree on algorithms. Step two consists of the exchange of the necessary cryptographic parameters to allow the client and server to agree on a premaster secret. Step three consists of the certificate exchange. In the fourth phase both the client and the server create a master secret. In the fifth step the security parameters are passed to the record layer. In the sixth and last step the client and the server verify that their peer has calculated the same security parameters and that the handshake has occurred without tampering by an attacker. The following sections will discuss each handshake phase in detail with regard to our assumed client-server scenario [3]. The negotiation phase is the first phase of a TLS handshake. The TLS handshake in our client-server scenario is initiated by the client. The client initiates the connection by sending a `ClientHello` message to the server. Note, that the negotiation phase could also have been initiated by the server with a `HelloRequest` message. The server must answer to a `ClientHello` message with a `ServerHello` message or else the connection will fail. These messages are used to establish the available capabilities of the client and the server. A `ClientHello` message consists of the highest supported TLS version, suggested cypher suites, suggested compression methods, a random value, supported extensions and, if the connection is resumed, a session ID. A cypher suite is a named combination of authentication, encryption, message authentication code and key exchange algorithms. Cipher suites will be explained in more detail in Section 8.6.2. Listing 8.11 shows an example of a `ClientHello` message.

```
struct {
    ProtocolVersion client_version;
    Random random;
    SessionID session_id;
    CipherSuite cipher_suites<2..2^16-2>;
    CompressionMethod compression_methods<1..2^8-1>;
    select (extensions_present) {
        case false:
            struct {};
        case true:
            Extension extensions<0..2^16-1>;
    };
} ClientHello;
```

Listing 8.11: ClientHello message [3].

The server answers to this message with a `ServerHello` message when it was able to find an acceptable set of algorithms. This means the server recognized some of the algorithms

in the cypher suites suggested by the client and is willing to use them. If it cannot find matching algorithms the server will answer with a handshake failure alert. The `ServerHello` message contains also a random value generated by the server. The clients random value and the servers random value are later used to generate a master secret from the pre-master secret. Listing 8.12 is an example of a `ServerHello` message.

```
struct {
    ProtocolVersion server_version;
    Random random;
    SessionID session_id;
    CipherSuite cipher_suite;
    CompressionMethod compression_method;
    select (extensions_present) {
        case false:
            struct {};
        case true:
            Extension extensions<0..2^16-1>;
    };
} ServerHello;
```

Listing 8.12: ServerHello message [3].

For the sake of simplicity this section will not discuss how extensions are handled or what extensions are available for the client and for the server [3].
After the client and the server have agreed on the connection details and a key exchange method requiring a certificate was selected, the server will send a message containing his certificate. The message will contain a certificate that is appropriate for the selected key exchange algorithm. Listing 8.13 shows the structure of a `ServerCertificate` message.

```
opaque ASN.1Cert<1..2^24-1>;

struct {
    ASN.1Cert certificate_list<0..2^24-1>;
} Certificate;
```

Listing 8.13: ServerCertificate message [3].

Depending on the scenario the server will send additional messages to support the key exchange. In this scenario it is assumed that the client and the server have agreed on using the RSA algorithm which does not require additional messages. If the server is done sending messages it will send a `ServerHelloDone` message. This message indicates that the server will not send any additional messages until he receives a message from the client again. After the reception of this message the client should verify that the server has provided a valid certificate and check if the provided server parameters are acceptable [3].
In our scenario it is assumed that the client and the server have agreed on using the RSA algorithm for key agreement and authentication. When the client has received a `ServerHelloDone` message it will begin to verify the certificate supplied by the server. This is done by checking the signature of the certificate to ensure that the certificate has not been altered. This is done by decrypting the signature on the certificate supplied by the server. The certificates signature is encrypted with the public key of the issuing certificate authority. If the client is not able to decrypt the signature with the public key of the issuing certificate authority something is wrong with the certificate and it should not be trusted. If the client is able to decrypt the certificate he can compare this value

against the hash value over the first part of the certificate which was generated using the algorithm specified inside the corresponding certificate field. If both values match, the client can be sure that the certificate has not been altered. Depending on the used configuration the client may check if the supplied certificate has been revoked using one of the certificate revocation methods mentioned in 8.3.3.
When the client is finished verifying the validity of the server certificate, it will generate a 48-byte long pre-master secret. This is only the case if RSA is used, if other algorithms are used, the procedure may be different. This pre-master secret is then encrypted with the public key of the server. The public key of the server is known to the client, because it is included in the certificate. See Section 8.4.1 for more details on the structure and fields of an X.509 Certificate. The client will send the pre-master secret in a `PreMasterSecret` message to the server. Listing 8.14 shows an example of a `PreMasterSecret` message. Note, that this message is specific to the use of the RSA algorithm and may looks different if an other algorithm like Diffie-Hellman is used [3].

```
struct {
    ProtocolVersion client_version;
    opaque random[46];
} PreMasterSecret;
```

Listing 8.14: PreMasterSecret message [3].

The client and the server then begin to calculate a master secret from the pre-master secret, the client random value and the server random value exchanged in the hello messages. The result of this computations is a 48 byte long master secret. Note, that this step can differ if an other algorithm than RSA is used [3].
When the master secret is generated, a `ChangeCipherSpec` message is sent by the client and the server. This message will notify the receiving party that subsequent records will be protected under the newly negotiated cipher specifications and the keys generated in the step before. This messages are used to ensure that both the client and the server have completed their time consuming public key operations before one side starts using encrypted messages. The reception of a `ChangeCipherSpec` message causes the receiver to instruct the record layer to immediately change the reading state of the connection. Likewise, sending of a `ChangeCipherSpec` message causes the sender to change the write state of the connection to the newly agreed state [3].
Immediately after a `ChangeCipherSpec` message has been sent, a `FinishedMessage` will be sent. The `FinishedMessage` is the first message encrypted under the newly negotiated conditions. If one partner receives a `FinishedMessage` it must immediately verify that it is correct. Listing 8.15 shows the structure of a `FinishedMessage`.

```
struct {
    opaque verify_data[verify_data_length];
} Finished;
```

Listing 8.15: FinishedMessage message [3].

The `FinishedMessage` contains verify data, a value that is created by using an agreed upon pseudo random function (PRF) with the master secret, a string depending on the sender of the message (”‘client finished”’ or ”‘server finished”’) and a hash over all previous handshake messages as seed. The client and the server can then decrypt each others `FinishedMessage` and verify that the encryption works. After both the client and the server have verified the state of the connection, the TLS handshake is completed and the application phase can begin using the defined keys [3].

8.6.2 Cipher Suites

A cipher suite is a named collection of authentication, encryption, message authentication code (MAC), and key exchange algorithms used to set up a Transport Layer Security (TLS) secured connection. The exact structures of cipher suites are defined in the same document as the protocol they are used for. The example from Section 8.6 discussed the RFC 5246 TLS 1.2 Protocol. In appendix C of RFC 5246 is a list of cipher suites that can be used in conjunction with the TLS protocol. Table 8.16 shows an extract of the TLS cipher suite list.

Cipher Suite	Key Exchange	Cipher	Mac
TLS_NULL_WITH_NULL_NULL	NULL	NULL	NULL
TLS_RSA_WITH_NULL_MD5	RSA	NULL	MD5
TLS_RSA_WITH_NULL_SHA	RSA	NULL	SHA
TLS_RSA_WITH_NULL_SHA256	RSA	NULL	SHA256
TLS_RSA_WITH_RC4_128_MD5	RSA	RC4_128	MD5
TLS_RSA_WITH_RC4_128_SHA	RSA	RC4_128	SHA
TLS_RSA_WITH_3DES_EDE_CBC_SHA	RSA	3DES_EDE_CBC	SHA
TLS_RSA_WITH_AES_128_CBC_SHA	RSA	AES_128_CBC	SHA
TLS_RSA_WITH_AES_256_CBC_SHA	RSA	AES_256_CBC	SHA

Table 8.16: Different Cipher Suites [3].

The key exchange algorithm is used to determine if and how the client and the server will authenticate during the handshake. The cipher algorithm is used for bulk encryption of message streams once the TLS handshake is finished. The message authentication code (MAC) is used to create a cryptographic hash of each block of the message stream [3].

8.7 Economic Aspects of Certification

This section provides an overview of different business models that can be seen in the market for digital certificates, issued by Certificate Authorities. Certificate Authorities (CA) can be grouped into three groups - Commercial CAs that charge money for certificates, free CAs that issue free certificates to anyone, and private CAs that are often used inside big corporations. The following sections will discuss each of those three groups in more detail.

8.7.1 Commercial Certificate Authorities

To fall in this group a Certificate Authority has to charge money for their certificates. For our distinction it does not matter how much the authority charges. The major commercial certificates authorities are Entrust, GeoTrust, Thawte and Verisign. [6].
In Switzerland SwissSign is also a major player in the commercial certificate market. SwissSign is a subsidiary of Swiss Post. SwissSign is a globally acknowledged Certificate Service Provider and offers a variety of SSL and email certificates as well as managed PKI solutions. [7]
The costs for a certificate from a commercial CA vary greatly and often involve some sort of service and/or warranty fees. Most certificate authorities offer more than just a simple certificate. It is often possible to buy consulting services or even an entire managed PKI solutions [8]. In order to get an actual certificate singed by a commercial CA you have to complete a process throughout which your identity or the identity of the machine the

certificate will be issued to will be verified. To purchase a certificate at SwissSign the registrator will have register with their web shop to first purchase a certificate license which can then be turned into an actual certificate. A certificate license entitles you to receive an actual certificate. Depending on the type of certificate you ordered the verification process will take from a few seconds to verify the provided email address up to a few days to verify a government issued ID. If a certificate for a company is requested it is necessary to send in a government issued ID for the company owner and in some cases financial statements of the company. [9]

8.7.2 Free Certificate Authorities

There are organizations, who offer certificates to anyone, who is willing to register with them. In this section will cover the organization CAcert.org as example. CAcert.org is an Australian non-profit organization that issues certificates to the public for free [16] CAcert's goal is to promote awareness and education on computer security trough the use of encryption. CAcert uses an assurance point system to verify it's users. If a member is new to the community, he will start with 0 points an can only request an un-assured client certificate. This certificate will expire after 6 months an can be used to sign emails. The only verification necessary to get such a certificate is to prove the ownership of the email address it is issued to, by answering to a special "ping" email. A member can gain assurance points by meeting with two or more verified CAcert members or by providing a government issued photo ID. With more points members are allowed to request other certificates, such as code signing certificates or server certificates.
It is worthwhile to note that certificates offered by Free CAs offer the exact same functionality as certificates offered by commercial CAs. There is no such thing as a difference in "production quality" for certificates. One of the main reasons why you have to pay for commercial certificates is the fact that commercial CAs are trusted and their root certificates are included in most of the standard web browsers. In addition to the higher level of trust, commercial CAs often offer a variety of value added services such as customer service and support or insurances with their certificates. Free CAs can be great for people who operate a private server and want to use a secured connection. Free CAs can also be useful in use cases where it is not important whether or not the root certificate of the issuing CA is distributed by default - for example in email communication with friends who all trust the issuing CA. However because the root certificate of a free CA is usually not distributed with the standard web browsers a certificate warning will be displayed if a user tries to establish a connection with a server that uses a certificate issued by a free CA.

8.7.3 Private Certificate Authorities

It is possible to sign and manage certificates by using only free software. This makes it possible for anyone with access to a computer to open its own certificate authority. Operating a private certificate authority can be beneficial for large enterprises that need to establish trusted connections between multiple entities of the same organization. Take for example a software development company that needs to ensure that only verified members of the company can contribute code to a shared code repository, or an office that manages access to network storage resources with certificates. A private certificate has several benefits over a commercial CA. A company running a private CA is capable of issuing certificates much faster than a commercial CA, because they can adapt the verification process to their own needs. A private CA can also help to keep the costs down while maintaining an acceptable level of security. Drawbacks of a private CA are that such a CA can only used by users within the organization. If external users want

to establish a secured connection that is secured with a certificate that is issued by a private CA, they have to add - and trust - the root certificate of this CA first. This inconvenience can be avoided partially by the use of cross certification between two or more organizations.

8.7.4 Choosing a Certificate Authority Type

When choosing a certificate authority trust and costs pay a major role. If you want to have a certificate that is trusted widely you will have to pay money for that trust. If the certificate is only used for a small group of people or for the sole purpose of encrypting a connection you can benefit from a certificate that is issued by a free certificate. Private certificate authorities take some sort of middle ground between the commercial CAs and the free CAs depending on how they are used. From a technical standpoint it is absolutely possible to manage thousands of user certificates with a private CA. However it is not necessarily given that all users will trust this CA. At this point the discussion can become almost philosophical - why should I trust one CA but not another only because one is in business longer than the other? Do I really know who I have accepted as third party trustee via the embedded certificates in my browser?

8.8 Summary

This report highlighted the significance of digital certificates and how they are used in the public key infrastructure. This report also showed how certificate Authorities are run from an economical standpoint. It is quite safe to state, that those mentioned CAs will not lose in relevance because the need need in digital certificates will further grow due to its possible use in the internet of things. In general the use of certificates in machine-to-machine or service-to-service constellations becomes more and more relevant. However there are multiple considerations to be made, one of them being the increasing costs. Even today the prices - especially for extended validation certificates in SSL/TLS communication - are quite high and therefore the trade-off between security and cost accordingly big. So new business/pricing models will be of high demand. This can already been seen to an extent with managed PKI solutions and it may go further in the direction to offer certificates as services, as flat rates or even as pay-per-use model. Aside from economic aspects, there are other aspects discussed and thought about such as the problem of centralized and institutionalized trust. For example, which authorities are really to be trusted? And why? One other concern is that the CAs are not actually authorities when it comes to the content of the certificates they issue. [19] Then there is the question how secure the whole process to getting a certificate is, especially with multiple CAs, sub-CAs, distinct RA's and VA's. In the Section 8.3.1 it was discussed who should be able to function as a CA with focus on the question what the governments role should be. Furthermore it was shown that there exist alternatives like the free CAs ones in Section 8.7.2 and that in some scenarios a private CA can be the way to go.
There are many aspects to consider and obstacles to overcome when it comes to the future of digital certificates and public key infrastructures - already today two integral parts of the internet. At this point the reader should have gained a basic understanding of not only digital certificates and their role in a PKI but also about the economic aspects of centralized trust and the questions that need to be addressed in the future.

Bibliography

[1] D. Copper, S.Santesson, S. Farrell, S.Boeyen, R. Housley, W. Polk, *Internet X.509 Public Key Infrastructure Certificate and Certificate Revocation List (CRL) Profile*, Internet Engineering Task Force (IETF), RFC 5280, May 2008. `https://tools.ietf.org/html/rfc5280`, visited 25.11.2015

[2] P. Yee *Updates to the Internet X.509 Public Key Infrastructure Certificate and Certificate Revocation List (CRL) Profile*, Internet Engineering Task Force (IETF), RFC 6818, January 2013. `https://tools.ietf.org/html/rfc6818`, visited 25.11.2015

[3] T. Dierks, E. Rescorla *The Transport Layer Security (TLS) Protocol Version 1.2*, Internet Engineering Task Force (IETF), RFC 5246, August 2008. `https://tools.ietf.org/html/rfc5246`, visited 14.12.2015

[4] S.Turner, T. Polk *Prohibiting Secure Sockets Layer (SSL) Version 2.0*, Internet Engineering Task Force (IETF), RFC 6176, March 2011. `https://tools.ietf.org/html/rfc6176`, visited 10.12.2015

[5] *Asymmetric Encryption* `http://ipseclab.eit.lth.se/tiki-index.php?page=2.%20Background` visited 09.12.2015

[6] W. Chou *Inside SSL: The Secure Sockets Layer Protocol*, IEEE, August 2002. `http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1046644`, visited 14.12.2015

[7] *About SwissSign*, `https://www.swisssign.com/en/about-swisssign`, visited 02.11.2015

[8] *SwissSign Products*, `https://www.swisssign.com/en/products`, visited 02.11.2015

[9] *SwissSign FAQ*, `https://www.swisssign.com/de/swisssign-faq/35`, visited 15.11.2015

[10] P. Zimmermann *Why I Wrote PGP*, June 1991, updated 1999. `https://www.philzimmermann.com/EN/essays/WhyIWrotePGP.html`, visited 4.11.2015

[11] P. Zimmermann, *PGP User's Guide, Volume I: Essential Topics*, October 1994. `http://www.pa.msu.edu/reference/pgpdoc1.html#section-4`, visited 14.12.2015

[12] Z. Yan, *Trust Modeling and Management in Digital Environments: From Social Concept to System Development*, Nokia Reasearch Center, Finland, 2010, p.104

[13] M. Myers, R. Ankney, A. Malpani, S. Galperin, C. Adams, *X.509 Internet Public Key Infrastructure, Online Certificate Status Protocol - OCSP*, Internet Engineering Task Force (IETF), RFC 2560, June 1999. `https://tools.ietf.org/html/rfc2560`, visited 14.12.2015

[14] S. Santesson, M. Myers, R. Ankney, A. Malpani, S. Galperin, C. Adams *X.509 Internet Public Key Infrastructure Online Certificate Status Protocol - OCSP* Internet Engineering Task Force (IETF), RFC 6990, June 2013. `https://tools.ietf.org/html/rfc6960`, visited 14.12.2015

[15] *SERIES X: DATA NETWORKS, OPEN SYSTEM COMMUNICATIONS AND SECURITY* , ITU-T, December 2012. `https://www.itu.int/rec/T-REC-X.509-201210-I/en`, visited 14.12.2015

[16] *CAcert Community Agreement*, CAcert. `http://www.cacert.org/policy/CAcertCommunityAgreement.html`, visited 14.12.2015

[17] J.R. Prims, *Interim Report - DigiNotar Certificate Authority breach*, Fox-IT IT Security, September 2011. `https://www.ncsc.nl/english/current-topics/news/interim-report-on-diginotar-digital-breach-published.html`, visited 14.12.2015

[18] Google Online Security Blog *An update on attempted man-in-the-middle attacks*, Google, August 2008. `https://googleonlinesecurity.blogspot.ch/2011/08/update-on-attempted-man-in-middle.html`, visited 15.11.2015

[19] C. Ellison, B. Schneier *Ten Risks of PKI: What You're not Being Told about Public Key Infrastructure*, Computer Security Journal Volume XVI, Number 1, 2000

[20] A. Stamos, S. Stender, *Web Service Security*,iSEC Partner, November 2005. `https://download.microsoft.com/download/8/d/6/8d608524-0763-48b5-840b-0ae446996a14/MS_WSS_Dec_05.pdf`, visited 14.12.2015

[21] S. Boyen, *White Paper X.509 Certification Path Validation*, January 27. 2004. `https://www.entrust.com/wp-content/uploads/2013/05/pathvalidation_wp.pdf`, Entrust, May 2013, visited 5.12.2015

[22] S. Boyen, *White Paper X.509 Certification Path Validation*, January 27. 2004. `https://www.entrust.com/wp-content/uploads/2013/05/pathvalidation_wp.pdf`, Entrust, May 2013, visited 5.12.2015