University of Zurich[UZH]

Burkhard Stiller, Corinna Schmitt, Radhika Garg, Thomas Bocek,
Daniel Dönni, Guilherme Machado (Eds.)

Internet Economics VIII

TECHNICAL REPORT – No. IFI-2014.01

February 2014

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

ifi

B. Stiller et al. (Eds.): Internet Economics VIII
Technical Report No. IFI-2014.01, February 2014
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 2013 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 VIII" discusses a number of selected topics in the area of Internet Economics. The first talk "In-network Aggregation Techniques in Wireless Sensor Networks" discusses the two existing in-network aggregation techniques *message aggregation* and *data aggregation*. Talk two on "Economics of Information Centric Networking" presents the status quo of cache deployment in the Internet, explores economic aspects of Information Centric Networking (ICN), and analyzes the incentives of the various stakeholders in ICN. Talk three on the "Impact of WebRTC (P2P in the Browser)" presents WebRTC and analyzes the impact it has on Network Address Translation (NAT). The fourth talk on "Economic Aspects of Network Neutrality" discusses economic aspects of traffic management in the Internet as is and four alternative scenarios. Talk five, "Bitcoins – Hype or Real Alternative", presents the novel digital currency Bitcoins and the current major players in the market. Finally, "A Success and Failure Factor Study of Peer-to-Peer File Sharing Systems" analyzes the factors that caused file sharing systems based on the Peer-to-Peer technology to become a success or failure.

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 Corinna Schmitt, Radhika Garg, Thomas Bocek, Daniel Dönni, Guilherme Machado, and Burkhard Stiller. In particular, many thanks are addressed to Daniel Dönni for his 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, February 2014

Contents

Chapter 1

In-Network Aggregation Techniques in Wireless Sensor Networks

Marc Gasser

A Wireless Sensor Network (WSN) consists of distributed autonomous sensors to monitor physical or environmental conditions. WSNs cooperatively pass their data through the network to a main location, where it is usually connected to the internet. WSNs are already developed, in industrial use and are becoming more and more popular. The practical examples are in the area of environment and health monitoring, for example the measurement of ocean temperatures, collecting information on patients' conditions, or management of critical industrial areas as oil containers. But also checking the concentration of chemicals and gases, warehouse management and supply chain monitoring are some current applications in use. Today's existing wireless infrastructures and the potential of WSNs lead to an increasing demand for pervasive computing. However, this field of research of heterogeneous networks in combination with today's heterogeneous wireless networks, has not yet achieved a widespread popularity. Especially when looking at today's smartphones, with many pre-installed sensors, that could be an obvious extension of existing WSNs. In addition there is an increasing popularity of wireless technologies as for example Bluetooth Low Energy (BLE), a technology that also enables hardware to send push notifications to mobile devices in about 70 meters proximity. These new preconditions of moving and scalable WSNs require a new use case of In-Network Aggregation techniques for heterogeneous networks. The motivation bases on the fact, that the advantages to combine a heterogeneous network with wireless sensors and BLE are quite obvious, but widespread applications in use are not known. The objective of this paper is to analyze existing WSNs, their In-Network Aggregation Techniques and the potential of using this technology in today's widespread heterogeneous networks consisting of smartphone-, WLAN- and other hardware. In this paper is discussed, how a setup of a real-world WSN business application could look like and if the increasing popularity of smartphone technologies supports the current development of the internet of things [4].

Contents

1.1 Introduction

Depending on the application there are several options of aggregation techniques, when sensors cooperatively pass their data through a network to a main location. In this paper three widespread mechanisms are discussed. There are disadvantages and advantages of the chosen data aggregation mechanisms Tiny Aggregation (TAG), Secure Information Aggregation (SIA) and Adaptive Application Independent Data Aggregation (AIDA). Then there will be presented an application of Wireless Sensor Networks on the example of the Swiss Federal Railways and public transport navigation. In this example application a Wireless Sensor Network can be extended with related technologies such as Bluetooth Low Energy (BLE) Beacons, WLAN, GPS and smartphones as a global sink. Such a heterogeneous network implies major limitations not only on energy consumption but also on security and usability. However, the motivation is the unused potential of WSNs when looking at today's heterogeneous infrastructures and at the use case mentioned before. There are a lot more use cases within a WNS as for example actively routing customers, in-store analytics, proximity marketing to contact-less payments. These additional use cases are not further discussed in this paper.

1.1.1 Wireless Sensor Networks

A Wireless Sensor Network (WSN) consists of spatially distributed autonomous sensors to monitor physical or environmental conditions, such as temperature, motion, compass, accelerometer, or pressure. The sensors cooperatively pass their data through the network to a main location, the sink, which connects to other networks. WSN are bi-directional, also enabling control of sensor activity or allows remote updates of the software. The development of Wireless Sensor Networks was motivated by military applications such as battlefield surveillance.
A Wireless Sensor Network consists of many sensor nodes, each equipped with a radio transceiver, a microprocessor, and a number of sensors. These nodes are capable of independently forming a network through which sensor readings can be propagated. Each node has an autonomous processing capacity and data can be processed as it is passed through the network. A wireless sensor node usually consists of the subsequent elements:

- Sensors
- Microprocessor
- Power unit
- Communication unit

In other words, a network of self-powered and moving sensor nodes for collectively sensing environmental data which is performing data aggregation has to be reliable, efficient and accurate. WSNs consist of clusters of devices using sensor technologies deployed in a specific area. The data is communicated wirelessly to a central system. The sensors consist of a processor with modest processing power, a few hundred kilobits of storage and an energy source. These factors lead to the corresponding constraints of energy, memory, computational speed, and communications bandwidth.
Given the limitations of the equipment and the very high demands with which the nodes must operate, algorithms and techniques must be designed to provide strong and efficient energy consumption. The design of the physical layer, communication technologies and the information coding still represent important challenges for the new WSN technology.

1.1.2 In-Network Aggregation Techniques

However, the goals of WSNs must be to conserve energy and bandwidth and due to the CPUs on board, one can move the integration and filtering of sensor data into the network itself. The in-network aggregation is required to process the user's query by allowing sensor readings to be aggregated by other nodes, because of forwarding raw information is too expensive and individual sensors readings are of limited use when for example the average room temperature has to be calculated.

1.2 Data and Message Aggregation

By forwarding the information of sensors, we differentiate between data and message aggregation. These are the two techniques of collecting the sensor data from the sensor node by using aggregation algorithms. Message aggregation does not involve pre-processing of the sensor data and data aggregation does process the information before forwarding to another node if possible. In Figure 1.1 it is shown how these processes look like on the example of maximum temperature via different sensors to the node.

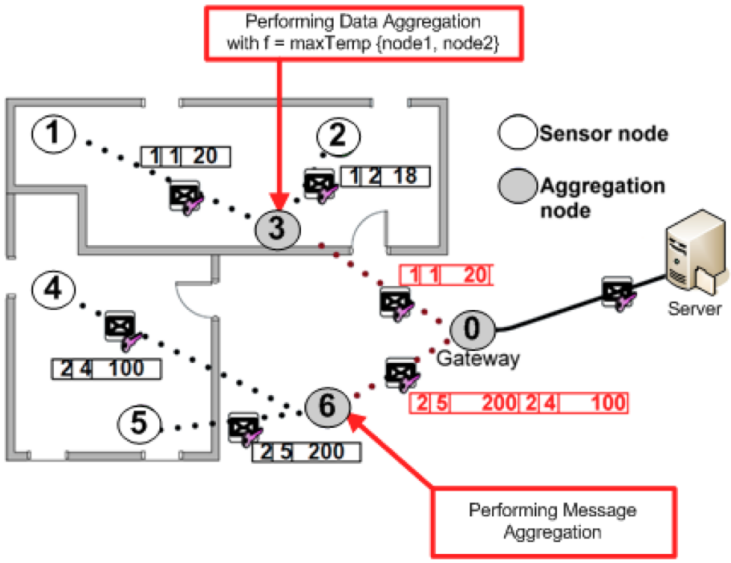

Figure 1.1: This Figure shows the difference between message and data aggregation, it shows that data aggregation has less data transmission than message aggregation [7].

1.2.1 Message Aggregation

Subsequent, Figure 1.2 is showing the energy consumption of a message aggregation which concatenates two or more messages into a newly generated aggregate message without data pre-processing. This is the standard algorithm and does not save energy compared to data aggregation algorithms.

1.2.2 Data Aggregation

Data aggregation implies lower energy consumption than simple message aggregation. It has to be considered that the calculation itself implies energy consumption as well, which is another reason why different kind of In-Network Aggregation Techniques have to be considered for each use case. In the example, by computing the aggregates the second step saves significant energy, depending on the requested types of values and aggregate

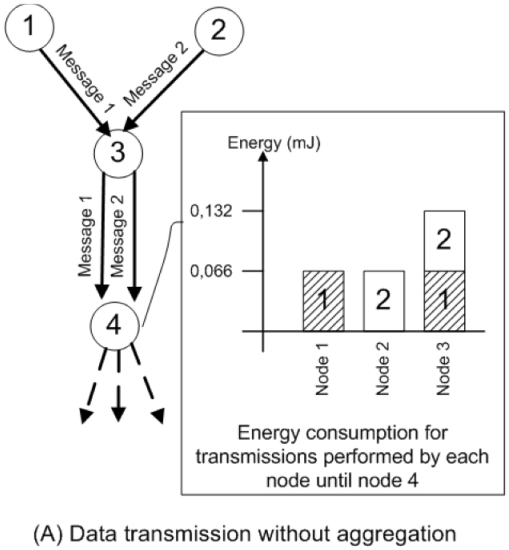

Figure 1.2: This Figure shows the general approach of message aggregation techniques [7].

functions - a median is most likely more demanding than a simple AVERAGE CALCULATION calculation [3]. But many sensor applications do not require individual sensor data and require aggregated data e.g. average temperature of the room. Aggregation in this setting refers to collecting data from various sensors and merging them in a single data point through aggregation functions e.g. SUM, AVERAGE, MAX etc.

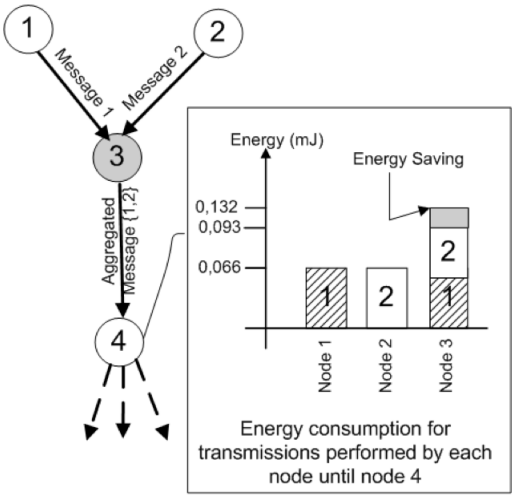

Figure 1.3: This Figure shows the general approach of data aggregation techniques [7].

1.3 Aggregation Techniques

In-Network Aggregation Techniques allow trading off communication for computational complexity. Given the use case, network resource constraints, and the fact that local computation often consumes significantly less energy than communication, In-Network Data Aggregation is the main topic of sensor network research. Resource efficiency, timely delivery of data to the sink node, and accuracy or granularity of the results are conflicting goals, and the optimal trade-off among them largely depends on the specific application. Initially, In-Network Aggregation Techniques involved different ways to route packets in order to combine data coming from different sources but directed toward the same destination. In other words, these protocols were simply routing algorithms that differed from more traditional ad hoc routing protocols to select the routing paths. Recently, additional studies have been published, addressing not only the routing problem but also mechanisms to represent and combine data more efficiently, taking care of security and scalability. It is a complex problem that involves different aspects of protocol design, and a characterization and comparison of all concepts and algorithms is missing in literature. Subsequent three examples of data aggregation algorithms will be introduced. The algorithms save total energy, compared to message aggregation techniques. The following algorithms focus on additional aspects as for example security or scalability. By transmitting a single bit of data is equivalent to many instructions. This communication dominates cost to send a message - many applications will benefit by processing the data inside the network rather than simply transmitting the sensor readings. The following three algorithms are an overview of how declarative aggregate queries can be distributed and efficiently executed over sensor networks. The in-network approach can provide a reduction in bandwidth consumption over approaches where data is aggregated and processed centrally.

1.3.1 Tiny Aggregation (TAG)

TAG processes aggregates in the network by computing over the data as it flows through the sensors, discarding irrelevant data and combining relevant readings into more compact records when possible. This is a benefit for a wide range of aggregate operations without having to modify low-level code or confront the difficulties of topology construction, data routing, loss tolerance, or distributed computing because of available query interfaces [3]. TAG is a tree-based approach and it is inspired by the selection and aggregation facilities in database query languages as for example MySQL, also the query interfaces are comparable. This distribution and collection phase is often designed for monitoring applications. The drawback is the inefficiency for dynamic topologies and it is very sensitive to link/device failures due to high energy consumption if the tree reorganizes.

1.3.2 Secure Information Aggregation (SIA)

A more secure algorithm is called Secure Information Aggregation and focuses on secure sensor networks that can handle malicious aggregators and sensor nodes. It is a demanding task to securely aggregate information in large sensor networks when the aggregators and some sensors may be malicious. SIA uses a simple aggregate-commit-prove framework for designing secure data aggregation protocols [6]. In Secure Information Aggregation aggregators or sensor may be attacked e.g. DDoS and implies that an aggregator reports different values. In the framework mentioned before, with the three phases aggregate, commit, prove it is possible to design a secure data aggregation protocol but there is a tradeoff between complex, very high security and efficiency.

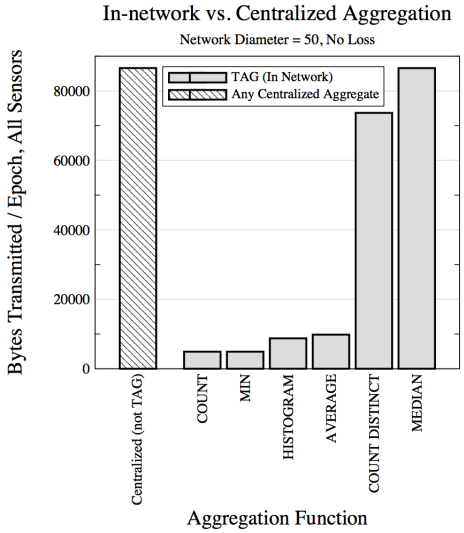

Figure 1.4: TAG data aggregation, [3]

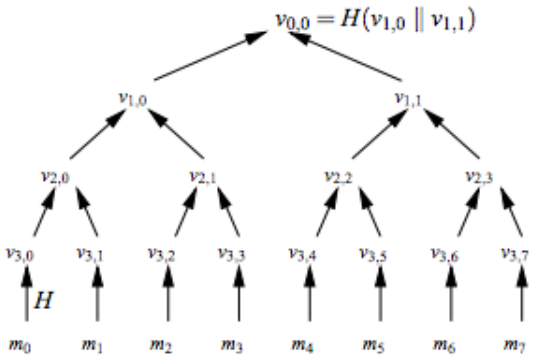

Figure 1.5: SIA data aggregation, general tree approach [6]

1.3.3 Adaptive Application Independent Data Aggr. (AIDA)

Adaptive Application Independent Data Aggregation is a mechanism for sensor networks which is more flexible for possible failures of certain nodes. AIDA takes the timely delivery of messages as well as protocol overhead into account to adaptively adjust aggregation strategies in accordance with assessed traffic conditions and expected sensor network requirements [2]. AIDA performs lossless aggregation by concatenating network units into larger payloads that are sent to first layer, the MAC layer for transmission. Due to the dynamic and quite often unpredictable nature of wireless communication in sensor networks, a innovative feedback-based scheduling scheme is proposed to dynamically adapt to changing patterns in traffic and congestion levels. AIDA is also a lossless aggregation technique.

1.3.4 Comparison of Data Aggregation Techniques

When comparing these data aggregation techniques there are individual advantages per use case which fit best to one aggregation technique presented. The adaptability in a wireless network e.g. when additional nodes become part of the WSN, AIDA masters this task best, where TAG and SIA have low adaptability features. When looking at error-proneness, of course AIDA is also here the best algorithm to be used, where for TAG and SIA an average energy consumption is needed. A big advantage of TAG is the scalability and efficiency in large WSNs over wide distances and many sensor nodes, where

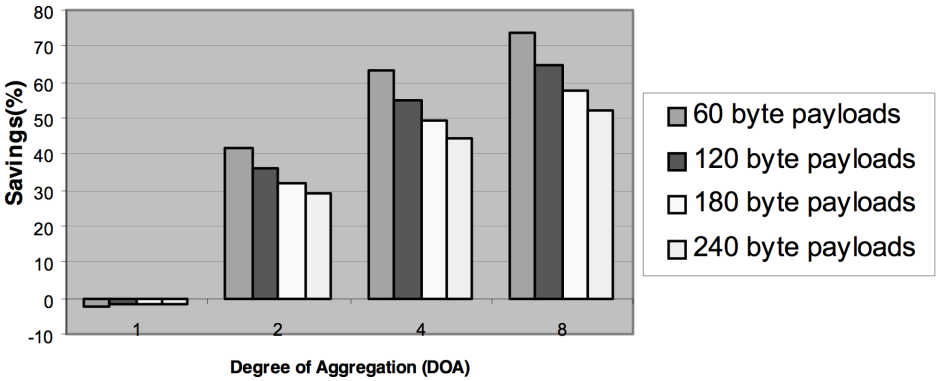

Figure 1.6: This figure shows AIDA and as the degree of aggregation increases, the percentage of savings in time increases drastically. [2]

SIA uses more energy than the AIDA for the security demands over wide distances and many nodes. Overall is the energy consumption lower for TAG than for SIA and AIDA when performing standard computations in big WSNs.
In the subsequent Table 1.1 the SIA technique is still highly efficient on a standard security level, this looks different when the security demands are a lot higher. TAG is a technique which has its advantages in the performance and scalability of bigger WSNs. And last but not least AIDA is fitting best if the sensors are running risk to fail while being exposed to different weather conditions for instance in an outdoor installation.

1.4 Protocols for WSN Communication

Before discussing an application of Wireless Sensor Networks on the example of the Swiss Federal Railways and public transport navigation, other wireless technologies shall be mentioned due to the fact, that a WSN can be extended with related technologies such as the new Bluetooth LE Beacons (which are more known as Apple's iBeacon technlogy), WLAN, GPS and smartphones as a global sink. After a short introduction an overview is presented in Table 1.2. In practical applications, the precondition of already installed e.g. wireless routers leads often to such heterogeneous networks [5]. In addition the new Bluetooth LE protocol is becoming popular, which is not comparable to the older Bluetooth protocols from v1.0 to v3.0 (cp. `http://en.wikipedia.org/wiki/Bluetooth`, 12, 2013). The LE standard is a different protocol than its predecessor which needed to be paired with a new device and often was not stable enough. "The specification for Bluetooth Smart, also known as Bluetooth LE, was released in June of 2010. Although bearing the Bluetooth name, it is a completely new specification, designed to enable very low-power devices ... Its history dates back to 2001, when Nokia attempted to have it selected as the core technology for the IEEE 802.15.4 initiative for a low-power, short-range radio. The IEEE group chose the proposal that was to become ZigBee, but the

	TAG	SIA	AIDA
Adaptibility	Low	Low	High
Error-proneness	Mid	Mid	Low
Scalability	High	Low	Mid
Security	High	Low	Mid
Efficiency	High	Low	Mid
Energy consumption	Low	Mid	Mid

Table 1.1: Comparison of three Data Aggregation Techniques

	WLAN	ZigBee, EEE 802.15.4	Bluetooth 1	BLE
Data Throughput	High	Low	Low	Low
Robustness	Mid	Mid	High	High
Range	50 - 300m	75m + mesh	50 - 1000m	10 - 300m
Large scale network	Very good	Very good	Bad	Very good
Low latency	Moderate	Bad	Very good	Very good
Pairing speed	Mid	Fast	Slow	Very fast
Power consumption	High	Very low	Mid	Very low
Cost	High	Low	Low	Low

Table 1.2: Comparison of Wireless Protocols

Nokia effort continued internally, eventually emerging as their Wibree" which influence Bluetooth LE later on, based on E. Vlugt, VP VeriFone about Bluetooth LE [8] p. 3 (see Table 1.2). Bluetooth LE has the main advantages of being highly energy efficient, at low cost and a very high pairing speed. Last but not least, smartphones have to be taken into account as well when discussing Bluetooth LE, the newer generation of smartphones is already equipped with Bluetooth LE and even switched on by factory default. The smartphone acts also as a node equipped with e.g. GPS sensors and can connect to every nodes via one connection to one node of the Wireless Sensor Network. Therefore, as soon as a device is connected, the entire Wireless Sensor Network knows where and which device connected to the network.

1.5 Heterogeneous Networks using BLE Beacons and WLAN

After understanding the advantages of TAG, SIA, AIDA and having an overview of wireless protocols, a project of the Swiss Federal Railways and public transport navigation at train stations shall be introduced to point out the need of a real-world use-case. The tender of this use case is publicly available and was already mentioned in several newspapers in Switzerland as for example in Tagesanzeiger (cp. `http://www.tagesanzeiger.ch/wirtschaft/unternehmen-und-konjunktur/SBB-planen-Navi-fuer-Bahnhoefe/story/14167642` 01.2014). This tender is a great example to understand how sensor nodes, cluster heads and sinks can be combined with Smartphones, GPS, Bluetooth LE Beacons and WLAN.
The Swiss Federal Railways will provide their customers the option to find the best route via their own mobile device when they arrive at the train station. The commuters would be guided when they arrive at the train station, by providing the best route to the track where the connecting train is leaving. This routing graph has to be calculated in real-time to be able to react on events as for example a broken escalator or a public event with high passenger streams. Another use case is to find specific Point of Interest in train stations, where a customer needs to be located as fast as possible via the mobile device.
The indoor positioning is one of the basic functionalities to show where a user is on a map [1]. The precondition that the Swiss Federal Railways installed several wireless routers in train station helps to cover this wide area by switching to the best available signals. For the positioning, the advantages of GPS, WLAN routers and Bluetooth LE Beacons can be used to have a more precise location information per device depending on what protocol is available at a certain location. This requirement implies using the existing WLAN routers and extending with Bluetooth LE Beacons to have a modest priced and stable infrastructure. A Bluetooth LE system can run for up to two years on a single coin

battery. After the setup, the system begins to transmitting 2.4 GHz Bluetooth signals. They can communicate with smartphones that are as close as a few centimeters away, or as far as 70 meters away. If a device can measure three or more signal strengths of Beacons, the position of the device can be triangulated with an accuracy of 1.5 meters. Apple with iBeacons, Google and Nokia have announced recently native support of BLE devices with their latest mobile operating systems and devices.
On this basis the routing of a person can be calculated on algorithms reflecting all sensors of the current infrastructure (e.g. broken escalators, people hotspots / clusters, weather conditions or events) and positions of the current commuters traffic. It is usual in WSNs to use non-deterministic communication channels characterized by variable delaying times when transmitting information. This makes synchronization in Wireless Sensor Networks the major task to be able to calculate in real time. In addition e.g. iPhones can not be used to triangulate via WLAN (Apple does restrict this information via WLAN) and should be positioned via Bluetooth LE where older Android phones can be tracked only via WLAN.
But also other opportunities come up by having an indoor positioning [1] which are connecting with the smartphone apps: from indoor analytics where to place a foodstore, proximity marketing based on the distance from a user to a product to buy, to contactless payments on the go. The Figure 1.7 shows what is possible with the Estimote (cp. `http://www.estimote.com` 01.2014) beacons. The demo is available to download for free (cp. `https://itunes.apple.com/us/app/estimote-virtual-beacon/id686915066?mt=8` 01.2014) and monitors distance but also has functionality to show relevant alerts based on movement. For this demo, one selects the beaconto monitor and move away from the BLE sender. One then is instructed to walk back towards the beacon as if entering a store. As one gets closer to the beacon, the message on the screen changes. This is possible, because the area around the beacon is divided logically into three zones (immediate, near and far), the same way they are visualized on the screen in the distance app. Whenever the user crosses the border between these zones, context changes and a meaningful message is appropriate.

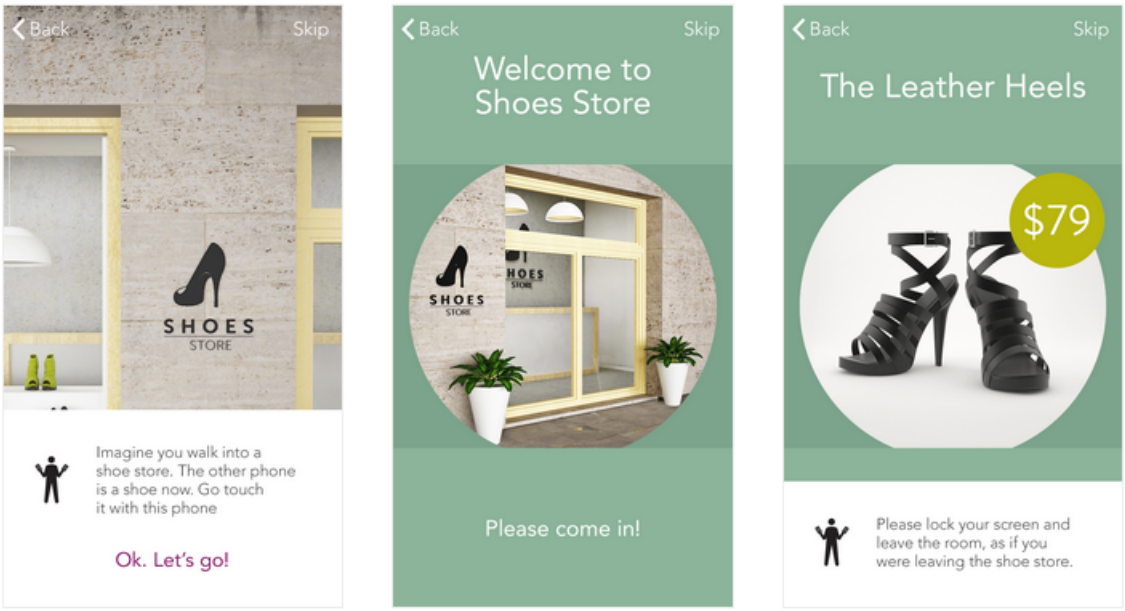

Figure 1.7: example estimote.com, proximity marketing based on distance between Bluetooth LE Beacon and smartphone

For those opportunities, the In-Network Aggregation Techniques come into place again. SIA for contactless and secure payment via mobile phones, TAG to combine the sensors

	distance sensor	outdoor sensor	moving sensor	payment
Data Throughput	TAG	-	-	-
Security	-	SIA	-	SIA
Large scale network	TAG	TAG	-	-
Pairing speed	-	-	AIDA	AIDA
Power consumption	TAG	-	AIDA	-

Table 1.3: Example setup of In-Network Aggreation Techniques with Wireless Protocols which are extended with sensor nodes.

over a long distance in e.g. the main station of Zurich. Then for sensors which are placed outdoors and are exposed to all weather conditions and could be damaged by people and a sensor could fail, AIDA. Also to locate service staff that is constantly on the move, AIDA could help to find the closest staff. For proximity marketing options as special offers when approaching a Point of Interest, push notifications can only be sent via Bluetooth LE to a wide range of smartphones. Figure 1.8 shows an example setup of a heterogeneous network.

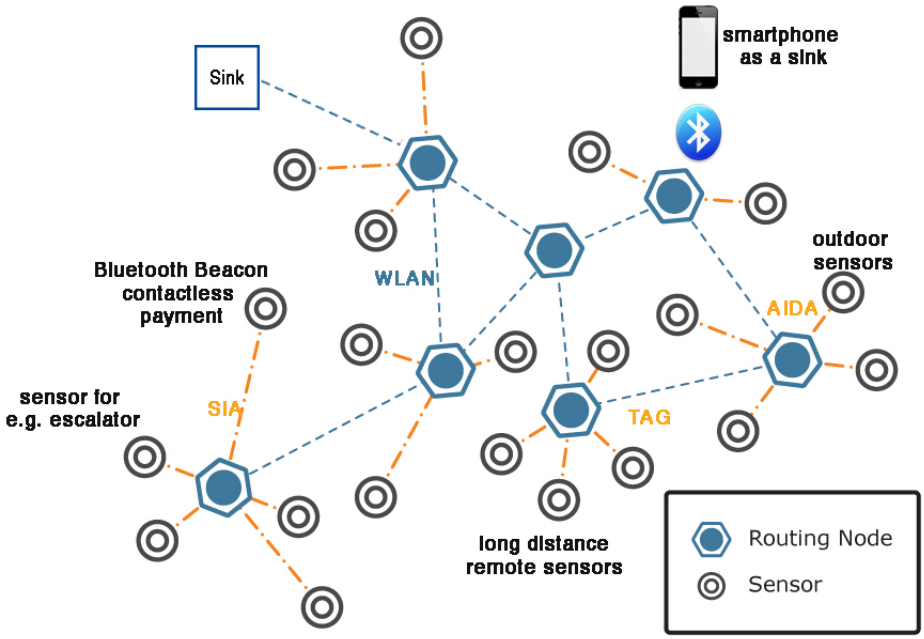

Figure 1.8: example of techniques and wireless protocols in a heterogeneous network

In general there are several attributes which fit more to a specific setup. If there is a need of sensors in a large scale network with high data throughput and low power consumption for instance TAG will be the most appropriate technique. For an outdoor sensor which could be protected from external access SIA or which to connect other far away sensors TAG. A moving sensor is often reliable on an own energy source and high pairing speed which AIDA could solve best. AIDA could also be appropriate for setup with a payment terminal or help when a sensor node will be extended with a smartphone to change the tree immediately. But if it comes to security the best option is still SIA.

1.6 Conclusion

In this paper, Wireless Sensor Networks were evaluated based on three kind of In-Network Aggregation Techniques. In addition a use case of a heterogeneous network on the example of Swiss Federal Railways was analyzed to show that the techniques discussed can fit in such situations.
The advantages of each mechanism are discussed. Compared to TAG, which fits best for long-distance and high performance needs, SIA gets less efficient in case of high complexity of the security algorithm. If sensitive data is gathered and the data doesn't have to be calculated in real time, still a good option for many applications. In a changing sensor environment AIDA is most applicable due its self-organizing capability if a sensor node fails.
The overview of wireless protocols gives alternatives to combine cluster nodes, where sensors usually transmit their data to the next node via the TCP/IP based wireless technology of sensor networks itself. The Bluetooth LE protocol has the advantage of being very stable, thus more appropriate for the fingerprinting or indoor positioning technique [1] since filtering algorithms are not necessary, as is the case with WLAN. Bluetooth is also widely available in computers and is a low-cost technology.
The example project of the Swiss Federal Railways and public transport navigation at train stations should encourage the thinking about other business cases about the topic of heterogeneous network in combination with smartphones and showed that the potential to combine Wireless Sensor Networks with related Technologies as for example Bluetooth LE Beacons and GPS. Such infrastructures as discussed can bring the richest possible customer experience to consumers. Last but not least, my learnings from this research:

- The potential of a Wireless Sensor Networks depends on the infrastructure which is available. There is no right or wrong, it is depending on the existing infrastructure and specific use case to decide which In-Network Aggregation Technique and which wireless protocols fits best and what configuration is needed most: Adaptability, error-proneness, scalability, security, efficiency, energy consumption, data throughput, robustness, range, latency, pairing speed and last but not least cost. If there is a need of sensors in a large scale network with high data throughput and low power consumption - TAG will be the most appropriate technique. For an outdoor sensor that needs to be protected from external access SIA and to connect other far away sensors TAG. A moving sensor is often reliable on an own energy source and high pairing speed which AIDA could solve best. AIDA could also help with a setup with a payment terminal for secure data transmission when a sensor node will be extended with a smartphone to change the tree immediately. But if it comes to security the best option is SIA.

- Major limitation in heterogeneous networks is not only energy consumption also security and usability. Security because sensitive data which could be collected by connecting to smartphones and contactless payment options and usability, when it comes to how easy is it to connect as a user and how compatible are the devices.

- BLE changes the paring behavior, it is not comparable to the older protocols from Bluetooth v1.0 to Bluetooth v3.0. It includes a range of broadcast advertising modes, which are essentially creating a much better user experience when pairing and connecting.

- Beacon applications are about to emerge in daily business. How much of its success will be hype remains to be seen. What is apparent is that the Apple effect, coupled with some innovative iBeacons app developers and the availability of iBeacons from many companies provides all of the ingredients for major innovation.

- There are technologies which might become obsolete in the discussed environment, for instance has NFC still a very high security level based on the low reach of an interactions but the key difference between the two technologies is that Bluetooth LE operates using a push strategy and the NFC using a pull strategy. When applied within proximity marketing campaigns, Bluetooth LE can fulfill all needs without NFC.

Bibliography

[1] Anja Bekkelien, Master Thesis *Indoor Positioning using Bluetooth* http://cui.unige.ch/~deriazm/masters/bekkelien/Bekkelien_Master_Thesis.pdf

[2] T. He, B. M. Blum, J. A. Stankovic, T. Abdelzaher, AIDA: *Adaptive Application Independent Data Aggregation in Wireless Sensor Networks*, ACM Transactions on Embedded Computing System, Special issue on Dynamically Adaptable Embedded Systems, Vol. 3, No. 2, 2004, pp 426-457.

[3] S. Madden, M. J. Franklin, J. M. Hellerstein, W. Hong *TAG: A Tiny Aggregation Service for Ad-Hoc Sensor Networks, SIGOPS Opereation Systems Review* Vol. 36, 2002, pp 131-146.

[4] McKinsey, *The Internet of Things* http://www.mckinsey.com/insights/high_tech_telecoms_internet/the_internet_of_things

[5] Rolf Nilsson, connectBlue *Digikey, Wireless Protocols* http://www.digikey.com/us/en/techzone/wireless/resources/articles/bluetooth-low-energy-for-wireless.html

[6] B. Przydatek, D. Song, A. Perrig, *SIA: Secure Information Aggregation in Sensor Networks, in: Proceedings of the 1st International Conference on Embedded Networked Sensor Systems* SenSys, ACM, 2003, pp 255-265.

[7] Dr. Corinna Schmitt *Secure Data Transmission in Wireless Sensor Networks* 3-937201-36-X, 2013, Dissertation

[8] Erik Vlugt, *Bluetooth Low Energy, Beacons and Retail* http://www.verifone.com/media/3603729/bluetooth-low-energy-Beacons-retail-wp.pdf2013

Chapter 2

Economics of Information-Centric Networks

Riccardo Patané, Julien Remond

Video traffic has been and will continue to be increasingly prevalent in the Internet. Streaming services are beginning of being deployed by the biggest video content providers. As the bit rate of multimedia traffic increases, the TCP/IP architecture may actually be inefficient for time-sensitive multimedia traffic delivery. Issues like multicast address assignment and complex group management also cause the endpoint-based Internet to be not suitable for multi or broadcast. Nowadays, peer-to-peer (P2P) systems deliver an important quantity of content to an big amount of users. But in general, P2P systems have limited information on the users downloading the content. From these observations and from the will to destroy the error 404, a new Internet architecture was born: Internet Centric Networking. In this article, we first introduce the ICN model and place it in today's context, with its advantages. We then have a closer look to the different stakeholders of the Internet and explain their potential incentives and breaks for introducing such an architecture. The next section of the paper will talk about the different costs that would be implied if the current architecture was changed for the ICN model, but also about the changes in the performance implied. We finally close with an overview of today's most controversy discussions about that model: the privacy and security problems, as well as the opposition of performance and costs.

Contents

2.1 Introduction and Motivation

Recently Information-centric Networking (ICN) has attracted significant attention in the scientific community. There has been various research initiatives, such as DONA [17], CCN [18], PSIRP/PURSUIT [6] [13], etc. with the aim of redesigning the future Internet, which can better satisfy the needs of today's and upcoming Internet traffic. When the Internet was created the communication requirements were very different than today. In that time the IP/TCP architecture was created for a trustworthy environment of fixed hosts that communicated over a dedicated network. Today the primary usage of Internet consists of content dissemination to a variety of PCs and mobile devices over unsecured multiple links. Thus, ICN research initiatives explore how we would design the Internet architecture if we would invent it today considering the current and future usage.
This paper examines some of the technical and economic aspects of ICN's. For this purpose, the literature that has been used for technical aspects is the CCN approach in [18]. For a more comprehensive overview of different ICN architectures the work of Ahlgren et al. [2] and Choi et al. [9] was really helpful. The economic analysis is mostly based on the work of Kostopoulos and Trossen [1] for analyzing Tussles and on the work of Agyapong and Sirbu [14] in order to understand the Incentives of the various possible stakeholders in an ICN.
The next section will give an overview of how ICN works and its differences in regards to today's Internet. In the second section, the main technical aspects of ICN are explained. The third part will examine the economics aspects of deploying caches: their cost factors, and the incentives and tussles of stakeholders in an Internet ecosystem.

2.2 Technical Aspects of ICN

In this section, after an introduction to ICN, the most important technical aspects will be explained in order to have a good idea about what ICN refers to and how it differs to current CDN model.

2.2.1 Introducing ICN

Information-centric networks (ICN) refers mainly to (i) a different way to get information from the Internet and (ii) a built-in caching function through which the information is delivered faster to end-users. The caching function is nowadays done by Content Delivery Networks (CDN). But rather than a network native function the CDN is the result of a technology that emerged out of necessity. The necessity for fast delivery of popular content intended to a big number of end-users. In other words CDN's could be considered as a patch for the inefficient TCP/IP architecture for mass content dissemination.
The different way to get information from the Internet means a paradigm switch from where to what. In the TCP/IP architecture there is a host-to-host communication. A connection between to endpoint (client-server) is first established (and authenticated) and then the information is transferred to the client. The ICN architecture does not care about the location of the information. It can be cached everywhere. The client just sends a request for a specific information to the Internet and the nearest device who has a copy of the data will serve the request by sending the information. This is done with a native caching functionality of the network itself. In order to provide this function efficiently the Internet needs ubiquitous storage infrastructure where information can be cached next to the end-users.

2.2.2 Main components of ICN

Current research topics in ICN to satisfy the above concepts are mainly following components of ICN architecture: Persistent Named Data Object (NDO), Naming, Caching, Application programming interface (API), Routing and Forwarding.

2.2.2.1 Persistent Named Data Object (NDO)

The idea is to put the data object in the center of attention and split the data in information pieces (objects, chunks) and to assign them unique and persistent names. This change of point of view from the where to the what allows finding the information by his name even if the data is moved from a location to another. The name of the chunk does not change over time. This implies that are no broken links, as the information is not bound to a location. In relation to ICN granularity refers to the chunk size. Rossi and Rossini give in [5] an overview of researches with different chunk sizes and their impact on the performance of ICN. Normally the chunk sizes are in the range of 1 KB to 10 KB.

2.2.2.2 Naming

There are different approaches for naming chunks. CCN describes a human readable, hierarchical naming structure (see Figures 2.1 and 2.2). It is similar to the current URL structure with a publisher prefix as the root of the hierarchy and where the / is the delimiter between the name components. The main advantage is the possible aggregation of the routing information and the improved scalability of the routing system [2]. A version marker (_v) captures the temporal evolution of the content and a segment marker (_s) identifies the chunks. Choi et al. [9] mentions that the URL like naming may imply a lower deployment hurdle due to the possible compatibility with the current URL based services and applications. It is also mentioned that semantics can prohibit persistent names. For instance if the name contains the information of the ownership and it would be at least misleading if the ownership changed.

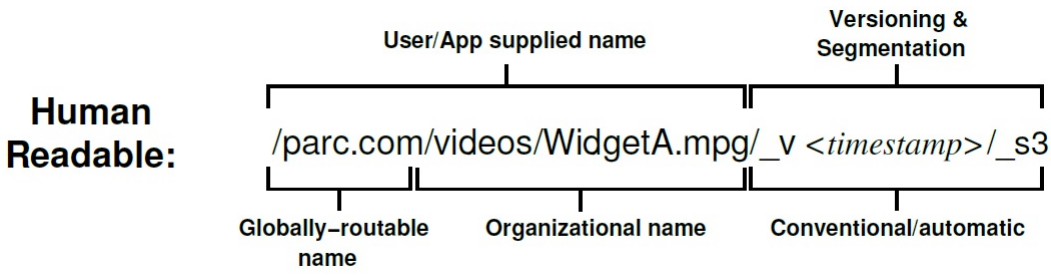

Figure 2.1: Example Data Name. [18]

2.2.2.3 Caching

Caching is an integral part of the ICN architecture. In ICN potentially all devices have caches, regardless of whether it is a node in the Internet Service Provider, (ISP) infrastructure or a computer in a home or corporate network. In addition, mobile devices are potential data caches. For instance, an information requester could get the information from a mobile phone of a person who sits next to her in an airplane.

Caching Performance Aspects. Since caching is also considered in today's CDN architectures, research in this topic is very extensive. However, the different architecture between CDN and ICN makes caching performance for ICN a widely unexplored research

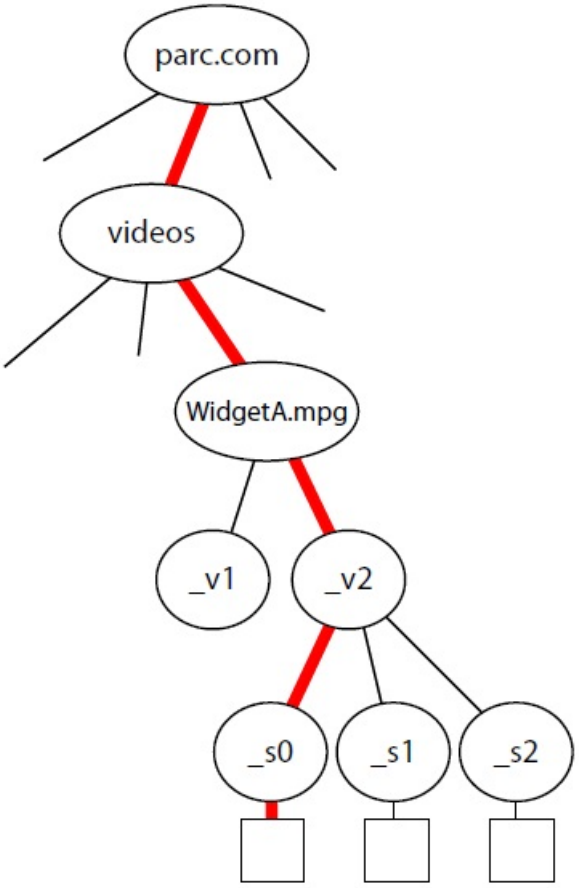

Figure 2.2: Hierarchical structure. [18]

topic. Caching performance aspects are highlighted in [5] by Rossi and Rossini. In the following sections a selection of the most important aspects are explained: cache and catalog size, content popularity, caching decision policies and replacement decision policies.

Cache and Catalog Size. According to Rossi and Rossini parameters describing the catalog are the chunk size, the object size, the router cache size and the number of objects of the whole catalog. The catalog refers to the sum of data available through the ICN. Psaras et al. defines in [8] cache as a relative factor. They associated the cache size with the traffic per second: cache is "the number of seconds worth of traffic cached in a given router". Table 2.1 shows some link speed and cache size combinations to determine cache.

LINK SPEED	1-SEC TRAFFIC	SECS OF TRAFFIC IN A 10GB CACHE
1,2 Gbps	~ 0.15 GBs	~ 64 secs
2,4 Gbps	~ 0.31 GBs	~ 32 secs
9,9 Gbps	~ 1.25 GBs	~ 4 secs
39,8 Gbps	~ 5 GBs	~ 2 secs
79,6 Gbps	~ 10 GBs	~ 1 sec
159,2 Gbps	~ 20 GBs	~ 0.5 secs

Table 2.1: Seconds of Traffic in a 10 GB cache. [7]

Rossi and Rossini [5] investigated the cache/catalog ratio of related work and estimated the ratio of YouTube and BitTorrent. As they could show there is a big difference between the ratio used in current literature and the real world. The current ICN literature is considering a ratio of 0.25% to 20%, whereas the real Internet had in the time of Rossi and Rossinis investigation a cache/catalog ratio seems more likely to be on the order of 10-5. Since Rossi and Rossini expect the performance of ICN been determined by this cache/catalog ratio, the benefits of ICN may be overestimated by current researcher.

Content Popularity. Perhaps the most important performance aspect is the content popularity. It is practically impossible to cache the data of the whole World Wide Web in every router. Therefore the ability of the routers to make caching decisions taking in account the popularity of the chunks received will impact the performance. Figure 2.3 shows the distribution of the number of views for videos on YouTube and Daum. We clearly can see a Pareto-like distribution indicating that 10% of the top popular videos register 80% of all views [10].

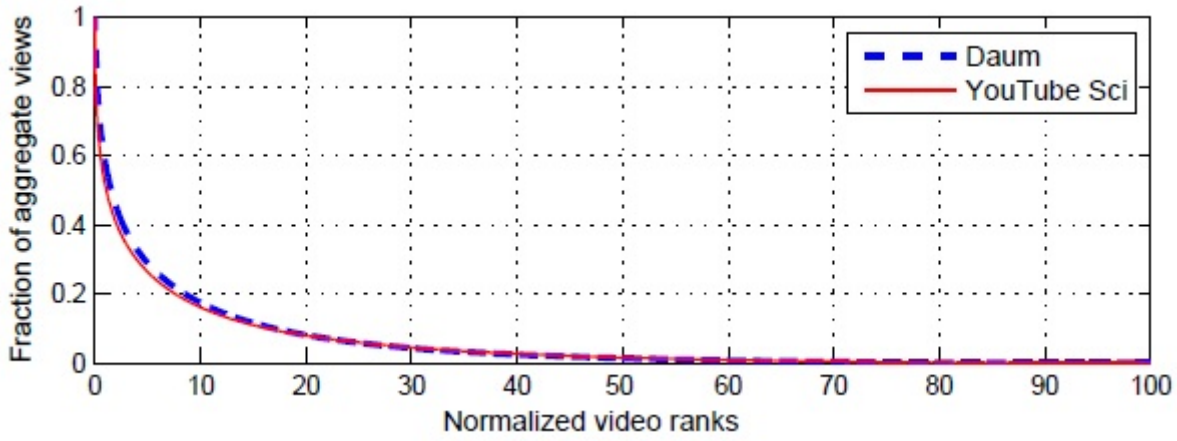

Figure 2.3: Skewness of user interests across videos. [10]

The shape of the curve in Figure 2.3 is in the Mandelbrot-Zipf law defined by the two factors: the skewness factor α and the plateau factor q. The higher the value of the q is, the flattened the curve will be. The reader is encouraged to consult [11] if he is interested on more information about the Mandelbrot-Zipf law. According to the experiments of Rossi and Rossini the setting of this two factors have a big impact on the overall system performance.

Caching Decision Policies. Decision Policies are rules to determine what incoming content hast to be cached. CCN [18] utilizes the Leave Copy Everywhere (LCE) approach. However, current literature proposes better policies like Leave Copy Down (LCD) and fixed probability (Fix(P)) [5], selective caching [15] and probabilistic in-network caching [8]. There is a trade-off between fast content delivery and reducing the content redundancy in an Autonomous System (AS). The mentioned policies aims to find the best compromise for this trade-off.
The LCD method will not distribute the content through the network at the first data request but will leave a copy on the path downwards to the client. The Fix(P) approach takes the decisions uniformly at random with a fixed probability between 0.75 and 0.9. Selective caching works with betweenness centrality in order to determine the best (the most central) node. In probabilistic in-network caching the probability to cache the content increases as closer as the cache is to the client.

Caching Replacement Policies. In opposite to the caching decision policies the replacement decision policies are those rules for deciding what data has to be deleted in the cache if there is no free space anymore in the memory for new content. Although replacement policies have also been long studied, Rossi and Rossini points out the fact that the data replacement has to be at line speed restricts the variety of viable policies for practice. They use the following replacement policies for their performance analysis: (i) Least Recently Used (LRU) replacement, (ii) First In First Out (FIFO) in which the oldest chunk is replaced, (iii) UNIF replaces a random chunk and (iiii) BIAS selects two random chunks and the most popular of them will be replaced. At the first look, BIAS surprises with deleting popular content. However, the assumption is that more popular content will be more replicated than less popular content.

2.2.2.4 Application programming interface (API)

This is all about commands and functions in order to request and deliver the NDO's over the IC-network. In CCN Interest Packets are used to request data and Data Packets are used to deliver the required information [18].

2.2.2.5 Routing and Forwarding

The CCN model, shown in Figure 2.4, suggests to use a Content Store for the Data Packets, a Forwarding Information Base (FIB) and a Pending Interest Table (PIT) in every cache enabled network device. In this context network devices encompasses all devices that are connected to the network, including Routers, Switches, Access Points, PC's, Mobile Phones, etc. The FIB contains the information of where to send the Interest Packet if the requested Object is not available in the Content Store. Hence, it works like a routing table of a common router. The PIT is a list of all incoming Interest Packets. This allows to gather all Interests for the exactly same data requested and if the data is not available in the content store, to forward the Interest only one time. Once the Interest could be satisfied, the Interest will be deleted from the PIT and the Data Packet is stored in the Content Store. If in a later time a further request for the same Object arrives, the Interest will be served from the Content Store and no Interest forwarding is needed.

2.2.3 Advantages of ICN

According to Ahlgren [2] the main advantages with an ICN architecture are:

- **Scalable and Cost-Efficient Content Distribution:** P2P file systems and CDN's have still problems like inefficient choice of peer or too much overhead. ICN tries to overcome this problems with a more architectural sound way.
- **Persistent and Unique Naming:** Today's network matches objects to locations with the help of DNS. As a result if objects are moved to an other location the object will not be reachable anymore. With the ICN approach data is found by their name and not by the location anymore. Thus, the data can be found regardless of it's location.
- **Security Model:** Security in the current Internet is obtained by trusting the server. For instance over the HTTPS protocol. Then the client assumes that the received information is the right one. ICN provides name-data integrity and origin verification and it is independent from which cache in the network the data was sent. Current Internet does not know this technique.

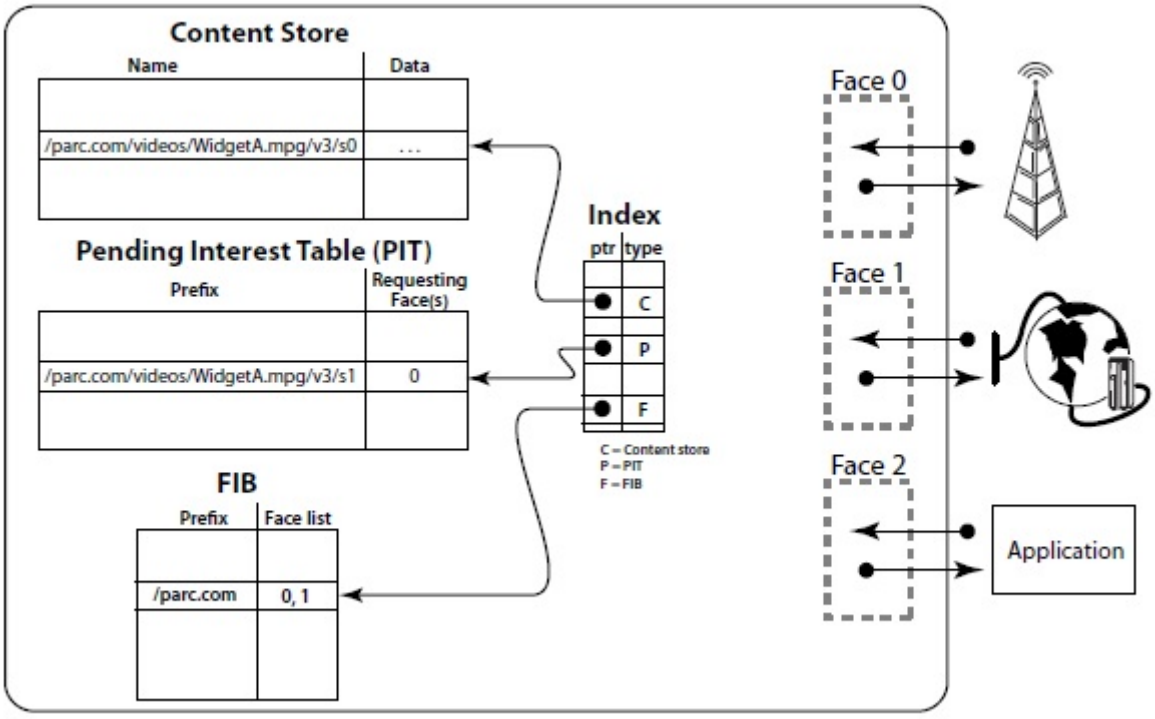

Figure 2.4: CCN forwarding engine model. [18]

- **Mobility and Multihoming:** Without end-to-end connections and with the ability to use multiple links to the Internet at the same time, the connectivity problem is easier to manage than today's network. Mobile clients just switch to new access and continue to send data requests.
- **Disruption tolerance:** In a challenged network a disruption in the connection can easily occur. If an application needs to bind the application protocol session to the transport session, the application fails as soon as the connection fails. The native in-network caching of ICN provides better reliability and performance by using a hop-to-hop approach instead of a end-to-end approach.

2.3 Economic Aspects of ICN

The ICN model seems to be the logical next step of today's Internet architecture considering nowadays' Internet usage [3]. But such a change of the caching network obviously implies a high cost and big economic repercussions. The following part of this paper will treat of that subject as follows: will first be introduced the stakeholders of the ICN model, the different incentives they may have for the model and finally the breaks they may encounter. The costs will then be discussed, both deploying and maintaining ones. Finally, one of the biggest debates generated by the model will be over viewed: Privacy and Security.

2.3.1 Stakeholders in ICN

Agyapong identified five different stakeholders in the ICN model [14]:

End-Users: End-Users are from the general market the persons that use the product. In the current context, they are the users of the ICN model, who access to content through the network.

Publisher: The publishers, or content providers, produce content for end-users.

Content Distribution Network (CDN): The Content Distribution Networks, or CDNs, primarily cache and deliver content on the behalf of the publishers in exchange for a fee.

Internet Service Provider (ISP): Internet Service Providers, or ISPs, provide network access to End-Users and Publishers.

Transit Network: Transit Networks are relay between the different Internet Service Providers (ISPs) to allow End-Users to access all the content of the Internet.
The Figure 2.5 illustrates one possibility of relationships between the different stakeholders in the model. The publisher and the CDN are connected to the ISP A while the End-Users have access to the network through the ISP B. Both ISPs A and B are connected via the Transit Network, allowing every stakeholder to be connected with one another.

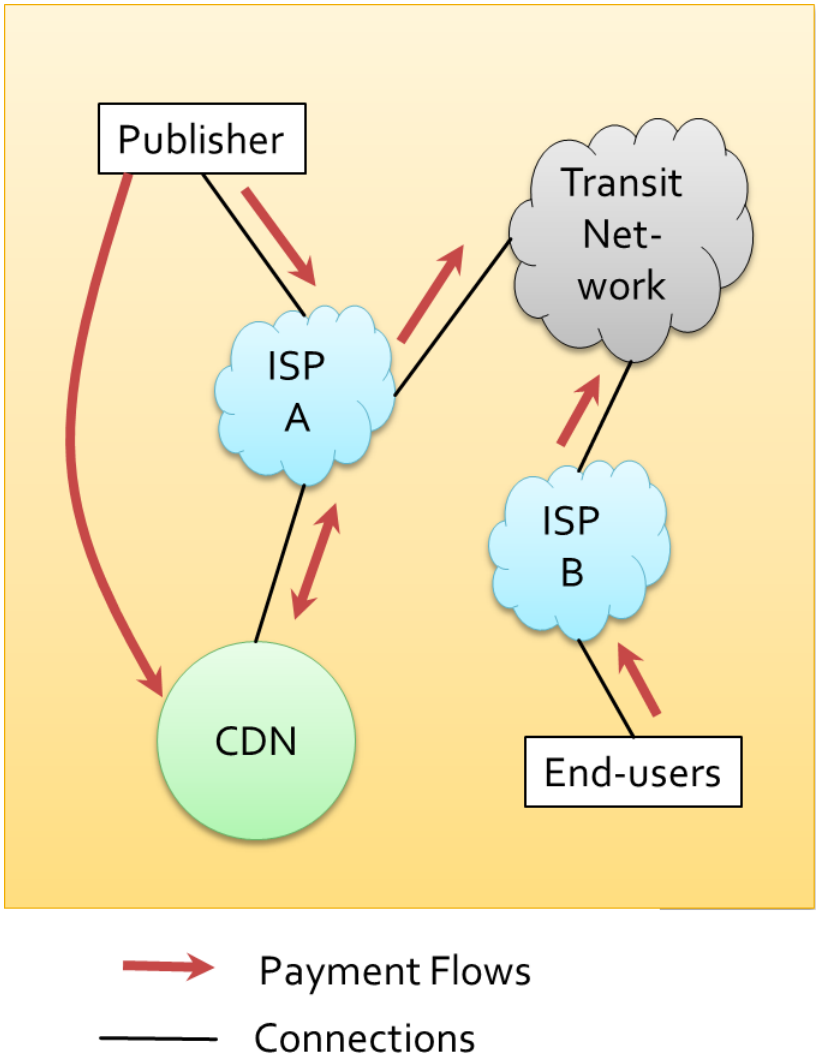

Figure 2.5: Possible Connections and payment flows.

2.3.2 Incentives

The key concepts of the model are expected to have an enormous impact on the network architecture in general, and will therefore create new opportunities for all the stakeholders.

According to Van Jacobson [18], ICN represents the third revolution in telecommunication networks as we move from connecting wires (public switched telephone network) to connecting nodes (all-IP networks) to connecting information (ICN).
There are three different types of Incentives that can be generated from ICN:

- Earnings
- Savings
- Technical Incentives

First, let's have an illustrated overview of the possible cash flow between all the stakeholders, meaning the first of the previous types: earnings. As you can see in Figure 2.5,

- ISPs receive money from publishers, End-Users, and some CDNs.
- CDNs receive money primarily from Publishers, but also sometimes from ISPs.
- Transit Networks receive money from the ISPs.
- Publishers indirectly receive money from the End-Users.

In addition of these earnings, some stakeholders may also be attracted to other aspects induced by ICN:

Publishers: providers of very popular content may face in the current network some traffic problems (in the one hand "Flash crowds", which is a high spike of traffic in a given moment, or in the other hand Distributed Denial of Service (DDoS) attacks). ICN would also allow content providers to do some savings: once the content is cached, there is no need to touch it anymore before the next update of content. Therefore, Network Access is only needed for updates, allowing publishers to make savings in Network Access [14].

Internet Service Providers: adopting the ICN model will generate a traffic reduction [14], and therefore important savings in Bandwidth. Indeed, popular content is nowadays only located at the Origin Server and at the Edge Servers (CDNs), causing for the hosting ISP a very high traffic. With ICN, content will be spread all over the globe, splitting the traffic among all ISPs reducing traffic for the previously named host.

End-Users: today, when an End-User sends a request to receive content, several steps happen: the DNS server sends back the IP address, the End-Users then connects to the closest Edge Server that may or may not have the asked content. With ICN though, once the request sent, the closest caches that have it will immediately send back the information. Assuming that these caches are closer to the End-User than the Edge Server of today's CDN architecture, End-Users will potentially get a lower latency from ICN.

Transit Networks: Transit Networks are the only ones among the stakeholders to actually have a lack of incentives for the ICN model [1]. Due to that architecture, information will be everywhere, and transit won't be as needed as it is today, causing a reduction of income for them. Transit Networks will therefore either have to increase their prices, or propose additional services, like content delivery (caching) to remain profitable.

2.3.3 Tussles

Stakeholders do have many incentives to ICN, but there also are many breaks that need to be worked on. These breaks are called Tussles. A tussle is a conflict of interest between two stakeholders generated by the ICN architecture. Five different categories of tussles are identified [1]:

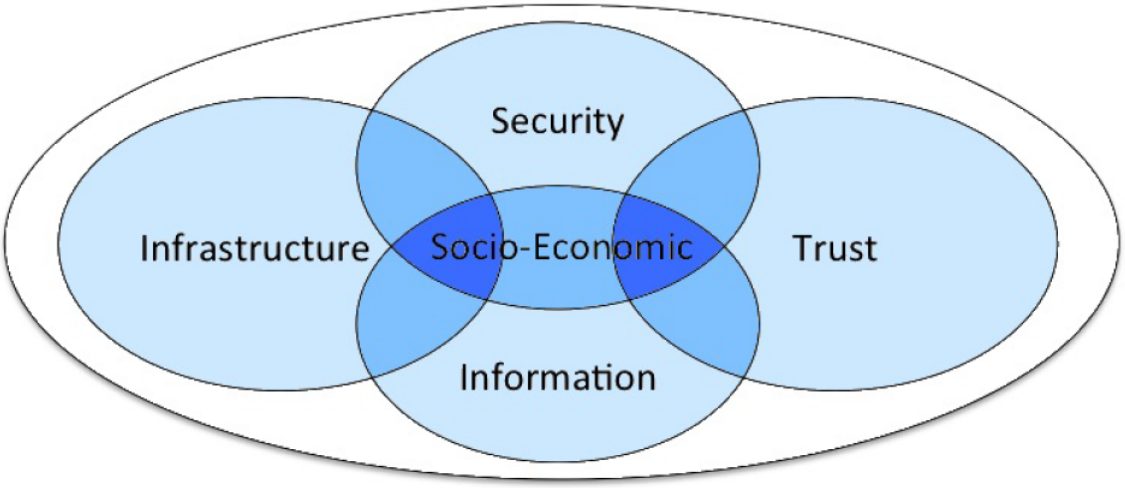

Figure 2.6: The different tussle categories. [1]

As can be seen in Figure 2.6, the categories of tussles are the following:

- Security: this category represents any possible security breach from one stakeholder to another.
- Trust: related to the previous one, the trust category characterizes any breach of trust between two stakeholders. Indeed, stakeholders sometimes have no choice but giving personal information to be granted access to some content.
- Information: this category mainly represents the management of the content itself, of the data gathered from end-users, etc.
- Infrastructure: these tussles concern the architecture of the networks.
- Socio-Economic: the socio-economic tussles are conflict of interest that suit to more than one of the previous categories.

To have a better understanding of these categories, the following will represent the identified tussle taxonomy. As can be seen, each tussle often involves more than only two stakeholders.

- Security

  Accountability: every End-User is responsible for its actions online. There is a conflict of interest between the privacy of these actions and the responsibility of requesting illegal content. (involved: ISPs vs. Regulators, End-Users vs. Regulators, Publishers vs. Regulators)

  Infrastructure security: when requested, the information follows a certain path: who defines these routes? (involved: ISPs vs. Publishers, ISPs vs. Regulators, Regulators vs. Publishers)

  Information security: End-Users often need to give information about them to have access to some content. Are these information secured? (involved: ISPs vs. End-Users, Publishers vs. End Users)

- Trust

  Trust in information: once the request for content is answered, and the content delivered, is that content the right one? Was it originally published by the original publisher? Is End-Users confidentiality granted? (involved:Publishers vs. Regulators, Publishers vs. End-Users)

  Trust in functions: is every function compliant to all regulations Are the misbehaving functions properly isolated/ corrected/ deleted? (involved: ISPs vs. Publishers, Publishers vs. Regulators)

- Information

  Brokering Information: conflict of interest between the will of the information holders and the information source: what are the regulations for profiling, matching interests and availability for spamming/advertising? (involved: Publishers vs. End-Users)

  Information Governance: who manages the information? Who owns the information space? (involved:Regulators vs. Publishers, Publishers vs. End-Users, ISPs vs. End-Users, CDNs vs. ISPs vs. Publishers)

- Infrastructure

  Delivering bits: requested information follows a predefined route to end to the End-Users. But sometimes, the requested content is in conflict with the regulations in application in the path. For example if someone in Singapore requests Facebook information from Japan, the information may need to pass through China, where Facebook is censored. (involved:ISPs vs. End-Users, Regulators vs. All the other)

  Brokering Topological Capabilities: in order to optimize the utilization of the network's infrastructure, its information needs to be exposed. (involved:ISPs vs. CDNs vs. Publishers)

- Socio-Economic

  Defining Functional boundaries: definition of boundaries for the execution of functions. Enforcement of these boundaries through technological, market and regulatory means. (involved: all the stakeholders)

2.3.4 Costs

As can be imagined, such a major change in the current Internet architecture is not cost free. The next step of the study is therefore the estimation of the costs in order to be able to define the worthiness of the project. Costs are divided in two different parts: the deploying costs, which is the initial investment to do to launch the new model. Will then be analyzed the maintaining costs, which are long-term expenses.

2.3.4.1 Deploying Costs

It is obvious that it is not yet possible to have an exact idea of the price of the patch. However, it is known that the first unavoidable and costly aspect of the change would be to have every router in the infrastructure to have a caching function. In order to do that, there are two possibilities:

- Change every existing router
- Use existing infrastructure by attaching CCN routers improved by caches.

Technology	Access time [ns]	Max. size	Cost [$/MB]	Power [W/MB]
TCAM	4	∼20 Mb	200	15
SRAM	0.45	∼210 Mb	27	0.12
RLDRAM	15	∼2 Gb	0.27	0.027
DRAM	55	∼10 GB	0.016	0.023
High-speed SSD	1,000	∼10 TB	0.03	0.00005
SSD	10,000	∼1 TB	0.003	0.00001

Table 2.2: Memory Technologies. [4]

A new ICN enabled router would cost approximately 1.5 to 2.5 times the price of a router that is used today.
As can be seen in the Table 2.2, prices really depend on the kind of technology the router will use. For example an edge router using the less expensive technology (Cisco 7505 Router with 1-TByte SSD for cache) will cost $24,000.00 for the router, plus $31,000.00 for enabling it for ICN.
→ The new router plus cache will therefore cost $55,000.00 in total.
A backbone router using a better technology (Cisco CRS 1 Series router with 80-GByte DRAM for cache) will cost $200,000.00 for the router, plus $130,000.00 for enabling it for ICN.
→ The new router plus cache will therefore cost $330,000.00 in total.
It is obvious that updating the current routers may be a much more viable option than replacing them all.
Let's now have a look at the different incentives of stakeholders for caching. Figure 2.7 illustrates the extent to which cache ownership allows a stakeholder to appropriate the monetary benefits from the value created in the network in absence of barriers to entry. We can see in Figure 2.7 a) that both Publishers and Eyeball Network (ISPs) share the benefits when the ISP deploys caches. However, the Publisher loses access to content delivery reports, and may want to pay for caching in order to obtain access to these reports [14].
When the ISP deploy the caching infrastructure and charge profit-maximizing prices, a deadweight loss is created for the system. In such a scenario, two constraints will limit the profit-maximizing price [14]:

- The ISP will raise the price until the deadweight loss is compensated.
- The profit-maximizing price will be bounded up by the price the lowest cost transit network will charge for caching if the ISP has no power in the termination charges negotiations [16].

This means that even though ISPs face significant barriers to entry or Publishers have significant transaction costs to pay in order to establish relationships with ISPs, it is probable that ISPs still have incentive to deploy limited transparent caches within their networks, as the ICN architecture is able to support this scenario by enabling ISPs to easily minimize duplications of efforts and required investments [14].

2.3.4.2 Maintaining Costs

Once the ICN architecture is installed, it becomes necessary to maintain it at an operational level. To do that, there will of course be costs involved.
Some of these costs are already presented in Figure 2.5, as they are very similar to the already existing ones. Indeed, the first expenses that maintain the Internet architecture

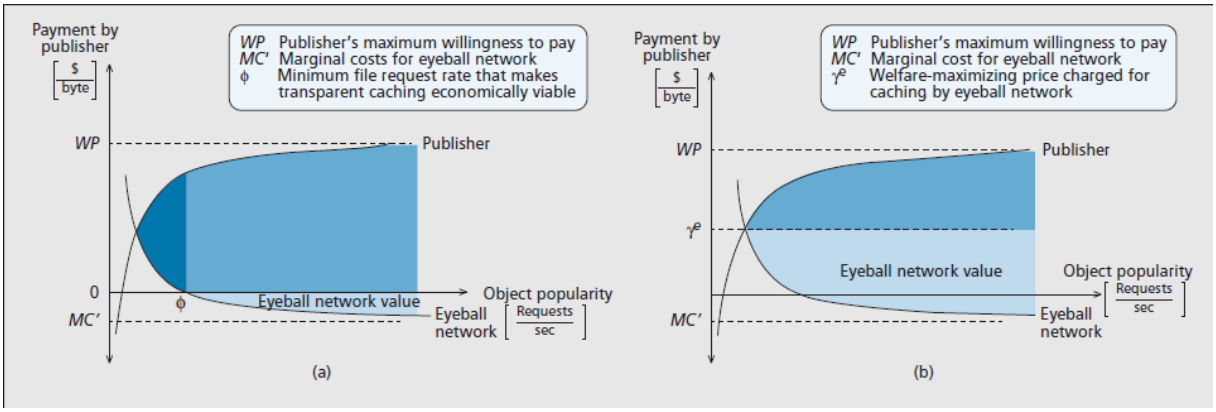

Figure 2.7: a) ISPs deploy transparent caches b) ISPs deploy commercial caches [14]

are the regular expenses among all stakeholders. But there are of course new costs that only apply in a model such as ICN.
A big change in the costs will be the security ones: publishers may have significant costs for signature generation or generation in order to be sure to authenticate the content and propose to end-user a trustworthy content [2]. The verification of the data on the routers is also a potentially important expense.
Table 2.1 shows the traffic speed of every different kind of link. Assuming an average cache size of 10-GBytes per router, we can see that a high priority link will spend 128 times less time in a router than a low priority one. Indiscriminate Universal Caching want that all content is cached with the same priority, meaning that most viewed content will be as cached as a viewer free content. Based on that table, we can see that caching YouTube, which counts 3 billion views every day, with the same priority as some random local news website will generate a higher need of verification of the data as the YouTube content will travel to slowly to be up to date. From that, we can say that Indiscriminate Universal Caching can be uselessly costly and suboptimal [7].

2.3.5 Privacy and Security

As discussed above, this architecture leads to some conflicts that may endanger both privacy and security of the final-users [1].
But most of the privacy issues are very similar to the one encountered today with the CDN architecture. Only further regulation can prevent the sell and usage of private information on the purpose of advertising.
Security, on the other hand, takes a new aspect in ICN. As the gathered information does come from its original creator, but it instead comes from some unidentifiable place, the information may be corrupted or outdated and therefore be either dangerous or useless for the final user. Fundamentally, ICN requires unique names for individual NDOs [2], since names are used for identifying objects independent of its location or container. It is important to establish a verifiable binding between the object and its name (name-data integrity) so that a receiver can be sure that the gathered bits actually represent the named object, and therefore be sure of the object authenticity. Associating the name of the object with its origin may be very useful.
A way to do that is to use the Public Key Infrastructure (PKI) authentication. The content is encrypted with a private ”lock” owned by the publisher. When published, the content is locked with that piece of encryption that is only known and can only be used by the publisher, ensuring that any content locked that way comes from this publisher. The publisher then spread a public key in the network that can be used by anyone to

unlock the content. Once unlocked, the content cannot be locked again, except of course by the original publisher.
The above function is fundamentally required for the ICN to be reliable. Otherwise, no one can trust the content's authenticity, which would enable several attacks, including critical Distributed Denial of Service (DDoS) attacks by injecting spoofed content into the network.

2.4 Performance vs. Cost

As can be seen, the ICN model can take form in a few different ways, and at different costs. But to be implemented for good, the model obviously has to offer both technical and monetary advantages to all or to a big majority of the stakeholders.
Here are the global impacts of ICN:

- Performance

  Low latency

  Low bandwidth usage

  Low traffic

  Low congestion

- Costs

  New Routers (with caching system)

  Higher maintenance costs

  Initial costs to learn technology and implement the model

As shown, ICN will allow stakeholders to face less traffic, to save bandwidth and to prevent some network congestion. End-Users will also potentially have a lower latency when asking for content. Most of the performance implications are computed 2.1, but a lot of them still depend on some choices. For example, what type of priority will be defined for content caching? In 2.3.4.2 is explained that a universal caching with no discrimination is pointless et very costly, but even if the content popularity system seems to be effective and the most cost efficient model, was shown in 2.3.4.2 that this system's viability depends on what fraction of the Internet's content is the most popular. Indeed, caching everything everywhere is not thinkable, and would be almost priceless. Regarding the implementation of the caches in ICN, is shown in 2.3.2 that ISPs do have incentives to deploy limited transparent caches within their networks. There are still discussions about conflict of interests among stakeholders, the tussles (2.3.3), and as said, many regulations will need to be implemented, costing both time and money, before considering implementing the architecture itself.

2.5 Summary and Conclusions

In this paper were presented all different aspects of ICN. Firstly, was introduced the ICN architecture in its technical aspects:its different components, the different caching strategies that can be used and their performances differences, and finally the overall performance implied by the model. Was shown that even if the model is only a patch to our current TCP/IP model, the ICN architecture is not easy to implement and requires a change of the whole caching system, as every router will need a caching system.

Has then be presented the economical aspects of the model: the different incentives and breaks for the model and its monetary implications. Was shown that almost every stakeholder has both technical and economical incentives the the ICN architecture. But the tussle taxonomy in 2.3.3 also shows the different conflicts of interest that would be implied by the model. That showed that the architecture still need a lot of work from the regulators to be considered as a viable option. Regarding the costs of the implementation of the system, they were proved to be very high, as adding a caching system to a router is very costly and it need to be done to every one of the existing routers, and more. But was showed that ISPs has the most incentives for caching in this environment.
Finally, were opposed these technical and economical aspects to provide an overview of the model and all its implications. It appeared that the model is not viable yet and that it may even be overestimated by current researchers.

Bibliography

[1] A. Kostopoulos and D. Trossen, "Techno-Economic aspects of Information-Centric Networking", Journal of Information Policy 2 (2012), 26-50.

[2] B. Ahlgren et al., "A Survey of Information-Centric Networking", IEEE Communications Magazine, July 2012

[3] B. Mathieu, J-F. Peltier, P. Truong, W. You, "Information-Centric Networking: a natural design for Social Network Applications". IEEE Communications Magazine, July 2012, 44-51.

[4] D. Perino and M. Vervello, "A reality Check for Content Centric Networking", ICN'11 Proceedings of the ACM SIGCOMM workshop on information-centric networking, p 44-49

[5] D. Rossi, G. Rossini. "Caching Performance of content centric networks under multi-path routing (and more)". Relatorio technico Telecom Paris Tech, 2011.

[6] D. Trossen et al., "Conceptual Architecture: Principles, Patterns and Sub-components Descriptions", May 2011. `http://www.fp7-pursuit.eu/PursuitWeb`

[7] G. Pavlou, "Information Centric Networking and In-Network cache management: Overview, Trends and Challenges", IFIP/IEEE CNSM 2013 Keynote speech.

[8] I. Psaras, W. K. Chai, G. Pavlou. "Probablistic in-network caching for information-centric networks". In Proceedings of the second edition of ICN workshop on Information-centric networking, ACM, pp. 55-60, 2012.

[9] J. Choi et al., "A Survey on Content-Oriented Networking for Efficient Content Delivery", IEEE Communications Magazine, March 2011

[10] M. Cha, H. Kwak, P. Rodriguez, Y. Ahn, and S. Moon, "I tube, you tube, everybody tubes: analyzing the world's largest user generated content video system", in ACM IMC, 2007, pp. 1-14.

[11] M. Hefeeda and O. Saleh, "Traffic modeling and proportional partial caching for Peer-to-Peer systems", IEEE/ACM ToN, vol. 16, no. 6, pp. 1447-1460, 2008.

[12] M. Sarela, K. Sollins and D. Trossen, "Arguments for an Information-Centric Internetworking Architecture". ACM SIGCOMM Computer COmmunication Review, Volume 40, number 2, April 2010, 27-33.

[13] P. Jokela, A. Zahemszky, C. Rothenberg, S. Arianfar and P. Nikander, "LIPSIN: Line Speed Publish/Subscribe Inter-networking", Proc. ACM SIGCOMM, Barcelona, Spain, 2009

[14] P. K. Agyapong and M. Sirbu, "Economic Incentives in Information-Centric Networking: Implications for Protocol Design and Public Policy", IEEE Communications Magazine, December 2013, 18-26.

[15] W. K. Chai, D.He, I. Psaras, G. Pavlou. "Cache less for more in information-centric networks". In NETWORKING 2012, pp. 27-40. Springer Berlin Heidelberg.

[16] S. Bauer, D. Clark, W. Lehr. "Interconnection in the Internet: The Policy Challenge", TPRC 2011: Proc. Telecommunication Policy Research Conference 2011.

[17] T. Koponen et al., "A Data-Oriented (and Beyond) Network Architecture", Proc. SIGCOMM '07, Kyoto, Japan, Aug. 27-31, 2007.

[18] V. Jacobson et al., "Networking Named Content", CoNEXT '09, New York, NY, 2009, pp. 1-12.

Chapter 3

Impact of WebRTC (P2P in the Browser)

Carol Alexandru

WebRTC, a technology that allows for real-time peer-to-peer communication from within a web browser, without any additional plug-ins, has been open-sourced by Google in 2011. Ever since, interest and support for WebRTC has been growing steadily. WebRTC has the potential to be a disruptive technology as it opens up a new realm for upcoming innovations that build on using the web browser for a variety of activities and purposes which had previously required specialized software: video and audio chat without the use of proprietary 3rd-party plug-ins, file sharing between peers without a server to relay the files, and multimedia conferencing without the need for proprietary, platform-dependent 3rd-party applications, just to name a few. WebRTC also impacts the commercial telecommunications market, affecting both OTT vendors and classical telco companies, as it enables more and more consumers to communicate exclusively over the internet, thus threatening the profitability of classical voice and text services. It is anticipated that by 2018, 4.7 billion devices will support WebRTC. There are yet a number of technical, political and economical problems with WebRTC that remain to be solved: peer-to-peer communication is difficult if the participants are located behind NAT and interoperability between different browsers must be guaranteed. A prominent topic of disagreement among the WebRTC community is the choice of 'mandatory-to-implement' codecs, which are subject to intellectual property laws. Given these open issues, the specification of WebRTC by the W3C and the IETF is still an ongoing effort. Nonetheless, WebRTC is already perceived as an influential factor in the future evolution of web technologies in particular and the real-time communications market in general.

Contents

3.1 Introduction

WebRTC allows web browsers to capture video and audio streams without the use of 3rd party plug-ins and to communicate with other WebRTC-enabled web browsers in a peer-to-peer fashion. An implementation of WebRTC has been open-sourced by Google in 2011 [3]. There are a number of motivating factors leading to the inception of WebRTC, many of them rooted in the desire to overcome technological barriers. This section first gives an overview over the structure of this report and the it presents the motivation behind WebRTC and the different actors who hold a stake in WebRTC.

3.1.1 Paper Outline

The remaining sections are organized as follows: Section 3.2 describes the architecture of WebRTC and the methods it uses to capture audio and video as well as to enable peer-to-peer communication within browsers, even if they are located behind NAT. Section 3.3 discusses the impact of WebRTC on different stakeholders and section 3.4 provides insights into some of the open problems that exist in the establishment of WebRTC as a standard. A final summary is given in section 3.5.

3.1.2 Existing Gaps in the 'Web Experience'

A classic web browser natively supports a certain set of features: Typically, it supports viewing rich text documents and images [21]. More recently, the HTML5 standard has added support for playing back video and audio sources, an accomplishment which had previously required the presence of third-party browser plug-ins such as Adobe Flash or Microsoft Silverlight [25]. Given this set of capabilities, the intended use for classic web browsers is to view remote content that is hosted on publicly reachable servers. Apart from PUT and POST actions (used for the occasional file upload or for entering small amounts of text), browsers are media consumers that are unable to produce any rich media. This topology limits the use of the browser and causes it to be unsuited in a number of potential scenarios: First and foremost, it is impossible to capture audio or video input from a microphone or a camera for the purpose of communicating either with the server or other clients. Various third-party vendors have developed browser plug-ins that add this functionality, but plug-ins are prone to security flaws [6, 30, 39] and may introduce compatibility and performance issues, or the plug-ins may be entirely unavailable on certain devices and platforms [32, 22]. Furthermore, the sharing of files between clients using only a web browser requires the presence of a server that receives the files from one client and then offers it to others [45].
WebRTC is an attempt at filling these existing gaps in the web browser experience by adding native support for capturing audio and video without the help of plug-ins and for communicating with other clients directly without the need to relay all the data through a server.

3.1.3 Lowering the Entry Barrier for RTC Technology

Products that make use of RTC technologies are traditionally difficult to develop. Developers need to consider many implementation aspects, spanning the domains of networking (connection establishment, authentication, encryption, etc.) and media processing (buffering, de-jittering, echo cancellation, noise reduction, encoding, decoding, etc.). They also need to find and license suitable codecs and ensure interoperability of existing products. These factors combined make the development of RTC technologies manpower intensive

and costly. Thus, the entry barrier for developers to create new RTC products is comparatively high. WebRTC attempts to provide a ready-to-use solution that implements most of the difficult implementation aspects. This not only reduces the load on developers specializing in RTC, but it also enables traditional web developers, who do not possess detailed RTC know-how, to experiment with RTC technology, in order to create new and innovative products [43, 33].

3.1.4 The WebRTC Ecosystem

The advent of WebRTC affects many stakeholders in various disciplines and the development of WebRTC and its accompanying standard involved many companies and organizational entities.
The main contributor to WebRTC development is Google, and Mozilla is also committed to the same WebRTC specification. Meanwhile, Microsoft is developing a competing standard, CU-RTC-Web, which shares many of the goals of WebRTC, but is fundamentally incompatible, as it does not use SDP for session descriptions [15]. Apple is almost entirely absent in the discussions around WebRTC [12], although they have started to attend W3C WebRTC Working Group meetings [31]. Behind the scenes, different interest groups participate in WebRTC development: The MPEG LA (Motion Picture Experts Group Licensing Authority) consists of many prominent companies such as Apple, Cisco, Microsoft, Hewlett-Packard, Sony, Ericsson and many more, which all hold patent rights to the H.264 video codec and other codecs. As such, the MPEG LA has a strong interest to promote its own codecs in order to collect royalties off the use of WebRTC. The W3C (World Wide Web Consortium) and the IETF (Internet Engineering Task Force) are attempting to specify the WebRTC standard, finding compromises considering both the interests of their corporate participants and the goal of an 'open web'.
Apart from the actors involved in creating and formalizing WebRTC, there are additional groups that play a role in the WebRTC ecosystem: Web Developers utilize the newly conceived capabilities of the browser to create innovative services and applications for both private and corporate users. Telecommunication firms are threatened by the move away from dedicated communication channels such as voice lines or SMS to an agnostic provider model, where audio, voice and text communication all occur over the internet. OTT (Over-The-Top) providers are competing with telco firms and to them, WebRTC is both a chance and a threat. Section 3.3 provides a deeper understanding of these different actors.

3.2 Technological Aspects of WebRTC

In May 2010, Google acquired Global IP Solutions (a.k.a. GIPS), a company that offered embedded solutions for transmitting audio and video data in real time, specifically for use in VoIP products, such as IP Phones or conferencing systems [10]. Reusing the work done by GIPS, Google developed a product called WebRTC, which they released as open-source in May 2011 [3]. Since then, integrating WebRTC in various browsers and other environments as well as the formulation of a W3C standard for WebRTC have both been an ongoing effort.
At the time of writing, recent versions of Chrome, Firefox and Opera support WebRTC by varying degrees. Since the W3C standard has not been fully formalized yet, browsers differ in their implementation and writing WebRTC-applications that run in all three browsers require browser-specific code. A Javascript library published by Google, adapter.js, provides a compatibility layer that smoothens the experience for the web developer by abstracting away browser-specific quirks [17].

This chapter gives an introduction to the architecture of WebRTC, highlighting the challenges that yet remain. It also touches the topics of security and privacy in the technical context of Web RTC and offers a glimpse into competing technologies that WebRTC is standing up against.

3.2.1 Architecture

WebRTC applications build upon a layered architecture: The *WebAPI*, which exposes the features offered by WebRTC, sits at the top. It is the sole interface used by web developers when writing Javascript applications that make use of WebRTC. Once the W3C standard has been formalized, the WebAPI should be the same in all browsers and environments that support WebRTC. Below the WebAPI lies the native implementation of WebRTC inside the browser. It is primarily up to the browser vendor how to structure the implementation and different vendors use different architectures [20, 37]. Nonetheless, they all have certain commonalities in that they need to provide the following functionality in one way or another:

- A transport component: In order to communicate with other clients in a peer-to-peer fashion, the browser needs to implement various protocols and procedures that enable it to connect to other peers and to exchange data. The transport component implements ICE (Interactive Connectivity Establishment), STUN (Session Traversal Utilities for NAT) and TURN (Traversal Using Relays around NAT), among other things, and it takes care of network I/O operations.

- A video component: The video component provides encoding and decoding procedures via the use of various codecs and it offers the functionality required to ensure a smooth stream, for example by employing jitter buffers and applying enhancements to the video stream. It enables the browser to stream video data, for example from a web cam or other video sources.

- An audio component: Parallel to the video component, the audio components offers codecs which are optimized for human speech, as well as additional features such as noise reduction and echo cancellation. It enables the browser to stream audio data, for example from a microphone.

Google has released a native implementation, libjingle, that enables browser vendors to provide WebRTC functionality by only implementing the adaption layer between the browser's own WebAPI and the native API, significantly reducing the amount of work required to add WebRTC functionality to an existing browser [19]. The native API can even be used in applications other than web browsers. For example, one could implement a voice calling app for a mobile phone without using a browser at all.

3.2.2 The Web API

To web application developers, the native API is invisible. They write applications in Javascript and make use of the Web API exposed by the web browser. Typically, this means working with three main components [7]:

- *MediaStream* (a.k.a. getUserMedia) allows the developer to receive data streams from a web cam or microphone. Some WebRTC-Implementations also allow streaming the screen output of the computer, for example for the purpose of screen sharing.

- *RTCPeerConnection* is responsible for streaming data between peers in a reliable and efficient manner. It incorporates features such as bandwidth adaptivity, dynamic jitter buffering, automatic gain control, echo cancellation for voice streams,

noise reduction and packet loss concealment. The developer uses RTCPeerConnection to establish connections, exchange session descriptions and perform signaling via a signaling server. These proceedings are discussed in section 3.2.3.

- *RTCDataChannel* can be used to augment the RTCPeerConnection to transmit not only media streams, but arbitrary data. The RTCDataChannel API is very similar to the HTML 5 WebSocket API, but all communication occurs peer-to-peer (whenever possible). RTCDataChannel can be used to develop arbitrary multi-peer applications, such as for example games, collaborative document editing or file transfer applications.

Using these three components, the application developer can compose rich peer-to-peer applications, combining media streams to enable video conferencing with data channels to implement additional functionality. One example for this combination is *CubeSlam*, a game developed by Google as a demonstration of WebRTC: The participants play a form of *Pong* in a 3D environment and each player sees a live video stream of their opponent on the other side of the playing field [18]. Another example is *Firepad Video*, a web application that allows multiple people to work collaboratively on source code and documents, and which incorporates a video chat [56].

3.2.3 Signaling and NAT Traversal

One of the primary goals of WebRTC is to enable Peer-To-Peer Connections between clients without relaying all the data through a server. However, servers are still necessary for two reasons: First, to serve the actual Javascript application that utilizes WebRTC and second, to initialize sessions between clients who wish to communicate. In order to establish a connection, clients need to exchange session descriptions (formulated using SDP [23]), which contain details on the form and nature of the data which shall be transmitted. This process is referred to as *Signaling*. Signaling in WebRTC is completely abstract, which means that there is no specification as to how it needs to be performed, as long as in the end, the session descriptions have been exchanged. This means that developers can use technologies for signaling which fit their problem domain. For example, a WebRTC application that involves VoIP softphones may want to use SIP for signaling. A chat application might choose to use XMPP and plain web applications might simply use HTTPS [9]. Figure 3.1 serves as an illustration for the typical use case: clients, which have loaded a web application from a server, perform signaling in a fashion chosen by the web application. They exchange SDP session descriptions and the native browser components establish a peer-to-peer connection between the clients to exchange media and other data.

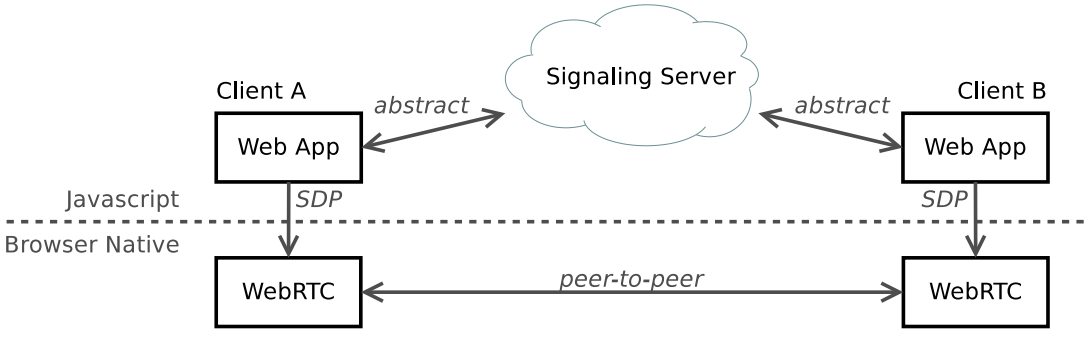

Figure 3.1: Two clients using a signaling server to establish a peer-to-peer connection.

After exchanging session descriptions, establishing the actual connection may involve NAT traversal which might be necessary if some of the clients are located behind NAT. NAT is a

mechanism that allows devices located on a private network to transparently communicate with devices located on an external network, such as the internet. It does so by modifying packets traveling between the networks, replacing the private source and destination IP addresses and ports with public ones where necessary while maintaining an internal lookup table of these mappings. In this fashion, devices behind NAT can communicate without even being aware that other participants may be located in different networks [51].
To understand how NAT traversal works in WebRTC, it is necessary to differentiate different types of NAT and to understand the different protocols that are involved in enabling NAT traversal, namely STUN, TURN and hole punching, which are used together in the ICE protocol. The following sections take a closer look at these individual topics.

3.2.3.1 Types of NAT

The next few paragraphs employ a hypothetical scenario, denoting IP addresses and ports by using variables, for example a:1 denoting a specific IP address a and a specific port 1. It involves the following entities:

- A client computer A behind NAT with a private IP address a
- A router R with a public IP address r
- Two STUN servers S1 and S2 with public IP addresses s1 and s2
- A public server P with a public IP address p

NAT traversal modifies packets as they pass through a router. If a:1234 sends a packet to p:80, R will modify the packet so that it appears to P as if the traffic was coming from r:x and not a:1234 where x is some port number chosen by R. This involves replacing all occurrences of a:1234 with r:x and remembering the mapping of private IP address and port to public IP address and port in a NAT forwarding table. Remembering the mapping is important, because the router will have to perform another replacement once the server answers. An incoming packet from the server P to the destination r:x will be edited so that the destination becomes a:1234. To both the client A and the server P, it seems like they are communicating directly, even though R performs network address translation. The exact procedure differs from device to device, but in general it is possible to discern four types of NAT [42]:

- Full-cone NAT is the least restrictive form of NAT. R creates a NAT forwarding table entry the first time A sends a packet to P. For example, the mapping may be (a:1234 -> r:x). Once this mapping has been established, any server Q can send data from q:z to r:x and R will forward the packets to a:1234, even if A never contacted Q.
- Address-restricted-cone NAT imposes a restriction on full-cone NAT, in that each port forwarding table entry is only valid for a specific target IP. For example, a mapping may be (a:1234 -> r:x to p) and only P can send packets to r:x. Packets sent from other IP addresses are dropped by R. The port is still unrestricted, meaning that P can send data to r:x from any port and the data will be forwarded to a:1234.
- Port-restricted-cone NAT imposes the additional restriction that only packets coming from the exact IP:port pair contacted by A will be forwarded. This means R may create a mapping (a:1234 -> r:x to p:80) when A contacted P on p:80, P can send data from p:80 to r:x for it to be forwarded to A. If the server P sends data on any port, then R drops the packets.

- Symmetric NAT is the most restrictive form of NAT. Same as with any other type of NAT, when A contacts P from a:1234, R chooses a port x at random and creates a NAT forwarding table entry, for example (a:1234 -> r:x to p:80). However, if A establishes another connection to P, even if using the same private port, i.e. sending data from a:1234, R will choose a second random public port y, creating a second mapping (a:1234 -> r:y to p:80). The public server P can reply to the packets coming from A via R, since it knows which port has been used (R replaced all occurrences of a:1234 with r:x or r:y inside the packet), but P only knows this if it received the packets coming from R. If P has not received any packets from R, P does not know the port which R has mapped for traffic to pass through to A. This condition has important implications for NAT traversal which will become clear in section 3.2.3.4

3.2.3.2 STUN

STUN is a protocol which can assist other applications in performing NAT traversal [49]. It does not perform NAT traversal by itself. The problem STUN solves is that a client behind NAT is unaware of its own public IP address and port, which are handled by the router. The client only knows its own private IP address and port. A STUN server is placed on the public web and clients behind NAT can send messages to the STUN server. The STUN server will then reply with a message containing the client's public IP address and port as seen from the STUN server. This means that STUN can be used by clients behind NAT to determine their public IP address and port. The client may even, to a certain degree, be able to determine the type of NAT employed by the router by sending requests to multiple distinct STUN server. When sending requests to different STUN servers using the same private port, the answers of the STUN servers will reflect the nature of the NAT employed by the router: For example, assuming a client A has a private IP address and port a:1 and uses this port to contact STUN servers S1 and S2, it may occur that both S1 and S2 reply to the request saying that the public IP address and port are r:9999. However, it may also happen that the two STUN servers give different replies, for example r:9999 and r:8888. The latter case would indicate that symmetric NAT is used by the router, since each outgoing connection has been assigned another port, even though A used the same private port to contact both STUN servers.

3.2.3.3 TURN

Another tool that may be used in NAT traversal is TURN [35]. TURN employs a publicly known server for clients behind NAT to connect to. In essence, two clients A and B both connect to the TURN server in a classical client-server topology and the TURN server then relays data between A and B. As such, if TURN is used, the traffic between A and B is no longer peer-to-peer. TURN can be used as a final fall-back if all other NAT traversal techniques fail, since normal client-server connections are possible in all NAT scenarios. However, relaying the data incurs additional cost. While STUN servers require very little resources to operate, TURN servers have high traffic demands, since all traffic between the peers is relayed.

3.2.3.4 Hole Punching

Hole punching is an umbrella term for a variety of techniques that can be used to perform NAT traversal between peers. The general mechanism by which it functions can be described as follows [14, 50]: If A and B are clients behind NAT routers R1 and R2, A and B first contact STUN servers to discover their public IP:port pairs. They subsequently exchange the pairs via a signaling server and then use the same private ports to contact each other. When A attempts to contact B, R2 will drop any incoming packet, because

R2's NAT forwarding table does not contain any entries that would allow A to contact B. However, in sending a packet to B, A has caused R1 to create a new mapping, for example (a:1234 -> r1:9999 to R2:9999), even though the packet was eventually dropped when it reached R2. Now, B can send a packet to r1:9999, causing R2 to create a mapping, for example (b:5678 -> r2:1111 to R1) as well and R1 will gladly forward the packet to a:1234, because this mapping exists in R1's NAT forwarding table. Now, both routers have created appropriate NAT forwarding table entries and peer-to-peer communication can take place.
This mechanism usually works for all types of NAT except when both parties are behind symmetric NAT. When replaying the scenario above with both A and B behind symmetric NAT, A would send a packet to B, causing R1 to create a mapping like (a:1234 -> r1:9999 to R2:9999) and R2 would drop the packet as usual. However, B will still not be able to send a packet to A because it cannot know which random port R1 has used to create the mapping, since it is almost certainly a different port other than the one reported by the STUN server in the first place. In principle, it is possible to traverse certain symmetric NAT configurations by contacting multiple echo servers many times to collect a large number of public IP:port candidate pairs. Given enough samples, it may be possible to predict the next port chosen by the router [55, 52]. However, there is no indication that this technique is used by current WebRTC capable browsers.

3.2.3.5 Putting it All Together Using ICE

WebRTC uses ICE, a protocol which utilizes STUN, TURN and hole punching, to establish a connection between peers. The procedure can be broken down into the following steps [47]:

1. Gathering candidates: As a first step, each client constructs a list of possible IP address and port pairs for other to connect to. These pairs are called 'candidates'. The first and obvious candidate is the local, private IP address and port. More pairs can be gathered by contacting one or more STUN servers, who will reply with public IP address and port pairs. Another candidate can be obtained by contacting a TURN server. At the end of the gathering process, the candidate list will contain several pairs.

2. Distributing candidates via the signaling server: The clients send their candidate lists to the signaling server which distributes them to the other clients. Now, all clients know the candidates of the other clients.

3. Connection and NAT traversal: Each client attempts to connect to other clients using the first candidate in the list. In case that a target client is not behind NAT, the first candidate - typically the IP address and port of the client machine itself, may already be sufficient. If the target client is behind NAT, the second candidate (returned by a STUN server and representing the public IP address and port of the router) may be the correct match. In case that the target client is behind a restricted NAT, hole punching is employed by the peers to establish a connection. Should this also fail, because all clients are behind symmetric NAT, the final candidate (representing the IP address and port of the publicly known TURN server) has to be used, and all traffic will be relayed via the TURN server.

By combining STUN, TURN and hole punching, ICE is able to penetrate most NAT scenarios, even those where all peers are behind symmetric NAT. If at least one client is not behind symmetric NAT, TURN is usually not necessary because hole punching can be employed in the direction of the client behind the non-symmetric NAT. In other words, the client behind non-symmetric NAT can send a packet to any client behind symmetric

NAT, thereby creating a new entry in the NAT forwarding table, after which the client behind symmetric NAT is able to contact the client behind non-symmetric NAT. TURN is only necessary when all participants are behind symmetric NAT.

3.2.4 Security and Privacy

WebRTC has been built with special regard given to security. All traffic is by default encrypted using AES in SRTP for media and DTLS for data transmission. It is in fact forbidden to use unencrypted RTP. When using HTTPS or another secure channel for signaling, WebRTC is secure end-to-end. For developers, this is especially beneficial since the biggest threat to encryption occurs when it is applied incorrectly and if all the security mechanisms are provided by WebRTC in the browser by default, there are fewer mistakes left for the individual developer to make [16]. The implementation has to be done correctly only once as part of the native WebRTC layer. From the usability side, WebRTC demands that browsers offer an explicit opt-in for all stream capturing actions. This means that the user has to explicitly allow the browser to capture the web cam, the microphone or the screen. The fact that WebRTC is implemented inside web browsers also means that it is likely that security holes will be patched rapidly, while traditional VoIP soft phones may not be updated for years [34].
A concern specific to WebRTC is the emergence of new social engineering techniques that may be used by malicious actors to access private information [26]. The classic scenario of an attacker calling an employee on the phone while pretending to be a 'service administrator' to extract passwords from the employee in order to gain access to the company infrastructure or steal secrets may be carried over to the realm of WebRTC, where a spy might be able to pose as just another participant in a video or chat conference where sensitive topics are being discussed.

3.2.5 Competing Standard

After Google had released WebRTC, Microsoft began working on a similar technology called CU-RTC-Web, which pursues the same goals as WebRTC, but which differs in a number of specification and implementation details. It was released January 2013 [15]. CU-RTC-Web provides a lower-level interface to networking and media components and it also avoids using SDP, instead providing a Javascript API for configuring sessions. The W3C WebRTC Working Group has decided to maintain the use of SDP even after Microsoft's proposal, but Microsoft is continuing development on their own standard [40].

3.3 Impact of WebRTC

Today, there are already over 1 billion endpoints in use which support WebRTC, such as computers equipped with Chrome, Firefox or Opera as well as Android mobile phones where both Chrome and Firefox support WebRTC [13]. It is estimated that this number grows to 4.7 billion by 2018 [2]. This section provides an overview on how different stakeholders are affected by the introduction of WebRTC and on the opportunities it offers as a driver for innovation.

3.3.1 Impact on Different Stakeholders

End users can probably expect a further decrease in cost for all forms of communication. Already, OTT-Providers, who offer text, voice and video communication over the internet, such as Skype, WhatsApp, Viber, are offering a free or relatively cheap alternative to

classic phone or SMS communication. This trend is likely to continue with more and more applications moving to the web [2, 11]. WebRTC, as a unifying method for writing peer-to-peer communication applications, also causes a decrease in diversity with regard to 3rd-party plug-ins and applications, such as Adobe Flash, Microsoft Silverlight or Java, which are currently necessary to fill the gaps in the web experience mentioned in chapter 3.1.2. On one hand, having fewer plug-ins means requiring fewer separate updates, increased security (because the code base to be maintained is smaller, more people are working on it and in many cases it is more transparent), and higher cross-platform compatibility of applications on multiple devices. On the other hand, multiple browser vendors all need to secure their product, which relativizes the security advantage of having fewer plug-ins, although vendors may share the implementation of the native WebRTC layer [34]. Furthermore, the introduction of richer communication comes with an increase in bandwidth and latency requirements. Users in areas with slow internet connectivity may not be able to profit from these new technologies to the full extent.
For corporate users, WebRTC represents an opportunity in various regards. First of all, contact with customers becomes easier. Services like the ones offered by Zingaya (a 'call us' button for websites which connects the visitor to a sales representative via WebRTC) simplify and intensify the communication between companies and customers [57]. Another opportunity arises from the reduction in cost and complexity when developing video conferencing applications. On one hand, corporate video conferencing is likely to become cheaper but on the other hand, existing investments in expensive, proprietary video conferencing solutions may face premature obsolescence [26].
WebRTC simplifies both the development and the distribution of various kinds of applications which previously required domain-specific knowledge. It is possible to write a video-chat application using WebRTC in less than 100 lines of Javascript. Already, there exist many innovative solutions, some of which are listed in chapter 3.3.2. In many ways, WebRTC is a prime opportunity of independent web developers and web startups to enter an emerging market [43, 33].
OTT vendors face a tradeoff in WebRTC. On one hand, OTT services become even easier to create, but on the other hand this threatens the position of existing OTT providers. Once more OTT services are utilizing WebRTC, it may become possible to connect different OTT services. This of course is beneficial for users, who will be able to communicate across OTT community boundaries, for example enabling communication between Skype and Viber users, but it means that the OTT vendor may lose the lock-in effect provided by the stand-alone application. OTT vendors will be forced to distinguish themselves through increased usability, service quality and value-added services [11].
Finally, the telecommunication faces the biggest threat from WebRTC. WebRTC intensifies the problem that OTT vendors have introduced, in that an internet connection becomes the only commodity that the user requires for all communication purposes. Some telco firms aspire to offer OTT services themselves, turning a threat into an opportunity. Some customers, especially businesses, may be willing to pay extra for higher quality of service internet connections, which prioritize WebRTC traffic to enable more reliable communication, but the implementation of these methods are non-trivial [11].

3.3.2 Innovations and Opportunities

For all stakeholders, WebRTC represents the advent of a new communication paradigm and a basis for innovation. The next few paragraphs list a small number of innovative examples that have emerged from WebRTC.
peerCDN offers a method for web hosts to reduce traffic to their servers. Instead of distributing data using a classic CDN (content distribution network), every visitor to a website becomes a peer and starts serving static files to other visitors. In this fashion,

visitors receive data not solely from the web sites web servers, but from other visitors as well. This reduces the load on the web server and shortens loading times for other visitors at the same time. It is also especially suited for handling traffic spikes, since more visitors to the site also means more peers that host static content [29].
Sharefest is another example for peer-to-peer file sharing using WebRTC. The user only loads the Sharefest web application from their servers and uses it to expose a local file on the web. As long as the user keeps the browser tab open, other users will be able to download the file via a URL directly from the owners machine. The file is not being uploaded to Sharefest servers, and once the user closes the tab, the file becomes unavailable [41].
Zingaya is a company which offers a 'Call us online' button, which companies can place on their website. Visitors can simply click on the button and be connected with a sales or support representative, without the need to pick up a phone, dial numbers and possibly even with video chat support [57].
But WebRTC not only provides room for innovation for web developers, since there is also an inherent need for infrastructure, even given that WebRTC works in the browser. There are a number of companies that offer WebRTC Signaling as a service, such as tokbox [54], XirSys [28] or TenHands [53]. These services make it easier for web developers to write WebRTC applications. There is also a need for TURN relays and gateway services that provide connectivity to POTS and SIP networks.

3.4 Open Issues

WebRTC has not yet been formulated into a standard which all stakeholders agree upon. At the time of writing, the three major browsers that support WebRTC, namely Chrome, Firefox and Opera, exhibit slightly different Web-APIs and some browsers does not support all the features envisioned by WebRTC [1]. Other popular browser vendors such as Apple and Microsoft do not support WebRTC at all and different stakeholders hold opposing views on some aspects of WebRTC [12].
Apart from the technical obstacles with regard to NAT traversal, a major discussion topic is the choice of mandatory-to-implement (MTI) codecs. In order for WebRTC to be successful, it is of the utmost importance that all WebRTC-enabled clients are equipped with at least one common codec, so that interoperability can be guaranteed. Clients may implement other codecs, for example to benefit from higher video quality, and use it for communicating with other clients that also support the same codec, but there has to be a fallback which can be used by all clients. While two MTI codecs have already been decided on for audio, namely Opus and G.711 [36], there has been no agreement on an MTI video codec. The two major contestants are H.264 and VP8. This section discusses the advantages and disadvantages of the two codecs as presented by their respective proponents.

3.4.1 Proponents of VP8

VP8 has been developed in 2008 by On2 Technologies, a company acquired by Google in 2010. Shortly after the acquisition, Google released the specification of VP8 under the Creative Commons Attribution 3.0 license and granted an irrevocable patent promise on its intellectual property rights on the implementation of VP8. In practice, this means that anyone can implement VP8 without having to pay royalties to Google. This fact is the main argument for VP8, as it allows any WebRTC client to implement and use VP8 without considering possible patent infringements. Followers of an 'open web' favor VP8 because of this freedom [4]

3.4.2 Proponents of H.264

On the other hand, the opponents of VP8 as an MTI codec present the following arguments, which have been formulated as an RFC [8, 48]: First of all, VP8 is not an established industry standard. Practically all chipsets on the market, be it for mobile or stationary communication devices, support encoding and decoding H.264, while only a small number of recent chipsets also support VP8. This means that devices which do not have hardware-accelerated VP8 coding will require the use of a software codec, which offers much lower performance and requires more energy. Many legacy devices, such as videophones used in corporate environments do not support VP8 at all. It is also unknown whether or not VP8 is indeed a safe harbor from patent infringements. In March 2013, Nokia filed a patent complaint with regard to VP8 [38], although the court decided in favor of Google in August 2013 [44]. It could yet happen that a company other than Google claims patent rights on some aspect of VP8. The legal framework around H.264 has been in place for many years and it can more safely be assumed that future patent disputes are unlikely. Proponents of H.264 also claim that given the young age of VP8, it simply is not mature enough, also noting that there are many H.264 implementations in use in countless applications, while there is only one VP8 implementation, which is only used in a small number of applications. Finally, H.264 proponents claim that H.264 High (the high quality profile for H.264) outperforms VP8 in terms of video quality [48].

3.4.3 Recent Developments

On October 30th 2013, Cisco (a member of the MPEG LA) announced that it will make available the source code and a binary package for H.264 starting January 1st 2014 [48, 46]. Cisco itself will pay the royalties to the MPEG LA for up to 100'000 distributed copies of the binary per year, per licensee. This is seen as a bold move of Cisco to push H.264 as the MTI codec. Nonetheless, the IETF was not able to reach an agreement for an MTI during the latest conference at the beginning of November 2013 [5]. There now is a discussion forming around the question of how to reach consensus in blocked discussions such as this, a topic which is in fact covered by IETF RFC 3929, since it is feared that the lack of consensus impedes the progress and proliferation of WebRTC [24, 27].

3.5 Conclusion

WebRTC has been envisioned as a solution for the problem that classical web technologies are unsuited for proper real-time communication, such as for example for real-time games, video and audio chat. Today, browser plug-ins are already being used to implement RTC solutions on the web, indicating a demand for the technology, but since plug-ins are prone to compatibility, security and other issues, WebRTC seeks to supersede them by providing a cross-platform and cross-browser solution. At the same time, WebRTC lowers the entry barrier for developers and innovators to join the market and create RTC products by encapsulating all the difficulties of implementing efficient and reliable real-time communication protocols. However, WebRTC is still an emerging technology: various, partially incompatible implementations exist and the formalization of a standard is currently hampered by arguments over technological and political details, such as the choice of MTI codecs.

Nonetheless, WebRTC is already being adopted eagerly by the web development community, with WebRTC applications already being used in production. Coupled with the optimism exhibited by analysts, bloggers and developers alike, this may indicate that WebRTC is more than just a hype phenomenon and that it is here to stay, whatever the final decisions on some of the specification details. OTT service and telecommunication

providers are aware of the threat that WebRTC represents. Realizing that fighting its adoption does not constitute a sustainable option for securing future revenue, they are advised to instead participate in the shaping and development of WebRTC.
From a technical perspective, one of the major obstacles for WebRTC use is symmetric NAT traversal. Since domestic users are one of the main target groups for WebRTC products and since many domestic routers employ symmetric NAT, finding solutions to the NAT traversal problem is an important task. There are a number of promising traversal techniques discussed in research and their future adoption by WebRTC implementations appears feasible. For the moment, TURN offers a largely reliable fall-back that allows WebRTC to spread even where symmetric NAT is employed, even though it incurs additional costs compared to the true peer-to-peer topology envisioned by WebRTC. It remains to be determined, whether or not WebRTC will truly provide better security and compatibility than the proprietary plug-ins and applications used today, but the open nature of WebRTC gives all stakeholders the chance to reflect on and improve the existing implementations. As WebRTC abstracts away many critical implementation aspects of real-time communication, web developers can focus on using the technology in higher-level use cases. Thus it acts as a driver for innovation, and solutions such as PeerCDN illustrate that WebRTC is being used for previously unanticipated purposes.
In conclusion, it can be said that WebRTC is a promising technology with broad demand and with strong support, both by companies and the web community. It may indeed have a significant impact on the web and telecommunication markets, forcing them to adapt and evolve, but also giving them a chance to grow and innovate. It is likely that activity surrounding WebRTC will increase further in the future.

Bibliography

[1] &yet: Is WebRTC ready yet? November 2013, http://iswebrtcreadyyet.com/ retrieved 20131127

[2] ABI Research: Future of Voice and Messaging - WebRTC and Telco APIs; September 2013, https://www.abiresearch.com/research/product/1014539-future-of-voice-and-messaging-webrtc-and-t/

[3] H. Alvestrand: Google release of WebRTC source code; Google, June 2011, http://lists.w3.org/Archives/Public/public-webrtc/2011May/0022.html retrieved 20131127

[4] H. Alvestrand, A. Grange: VP8 as RTCWEB Mandatory to Implement draft-alvestrand-rtcweb-vp8-02; IETF Network Working Group, October 2013, http://tools.ietf.org/html/draft-alvestrand-rtcweb-vp8-02

[5] V. P. Avila, C. Hart: WebRTC Video Codec Decision is...NO DECISION; webrtcH4cKS, November 2013, http://webrtchacks.com/ietf-finally-made-decision-mandatory-implement-mti-video-codec-webrtc/ retrieved 20131127

[6] J. Bau, F. Wang, E. Bursztein, P. Mutchler, J. C. Mitchell: Vulnerability Factors in New Web Applications: Audit Tools, Developer Selection & Languages; Technical Report, Stanford University, February 2013, http://seclab.stanford.edu/websec/scannerPaper.pdf

[7] A. Bergkvist, D. C. Burnett, C. Jennings, A. Narayanan (eds.): WebRTC 1.0: Real-time Communication Between Browsers (Working Draft) W3C, September 2013, http://www.w3.org/TR/webrtc/

[8] B. Burman, M. Isomaki, B. Aboba, G. Martin-Cocher, G. Mandyam, X. Marjou, C. Jennings, J. Rosenberg, D. Singer: H.264 as Mandatory to Implement Video Codec for WebRTC draft-burman-rtcweb-h264-proposal-03; IETF RTCWEB Working Group, October 2013, http://tools.ietf.org/html/draft-burman-rtcweb-h264-proposal-03

[9] L. Byrd: The Unbearable Lightness of WebRTC Signaling, WebRTC World, August 2013, http://www.webrtcworld.com/topics/from-the-experts/articles/350460-unbearable-lightness-webrtc-signaling.htm retrieved 20131130

[10] J. Dagenborg, W. Moskwa, M. Nesbit: Google buys Norwegian audio-video tech provider; Reuters, May 2010, http://www.reuters.com/article/2010/05/18/us-globalipsolutions-google-idUSTRE64H1O820100518 retrieved 20131127

[11] Disruptive Analysis: WebRTC Market Status & Forecasts: The hype is justified: it will change telecoms; February 2013, http://disruptive-analysis.com/webrtc.htm retrieved 20131130

[12] F. Donovan: Microsoft, Apple undercut mobile WebRTC development; FierceMobile IT, September 2013, http://www.fiercemobileit.com/story/microsoft-apple-undercut-mobile-webrtc-development/2013-09-26 retrieved 20131130

[13] S. Dutton: WebRTC Plugin-free realtime communication; Google Developers, October 2013, http://gotocon.com/dl/goto-aar-2013/slides/SamDutton_RealtimeCommunicationWithWebRTC.pdf

[14] B. Ford, P. Srisuresh: Peer-to-Peer Communication Across Network Address Translators; Proceedings of the USENIX Annual Technical Conference (ATEC'05), April 2005, http://dl.acm.org/citation.cfm?id=1247373

[15] A. Foresti: MS Open Tech publishes HTML5 Labs prototype of a Customizable, Ubiquitous Real Time Communication over the Web API proposal; Microsoft Open Technologies, January 2013, http://blogs.msdn.com/b/interoperability/archive/2013/07/24/new-cu-rtc-web-prototype-from-ms-open-tech-demonstrates-webrtc-video-support-without-sdp-offer-answer.aspx

[16] M. Georgiev, S. Iyengar, S. Jana, R. Anubhai, D. Boneh, V. Shmatikov: The Most Dangerous Code in the World: Validating SSL Certificates in Non-Browser Software; In the proceedings of the 19th ACM Conference on Computer and Communications Security (CCS'12), October 2012, p38-49 https://www.cs.utexas.edu/~shmat/shmat_ccs12.pdf

[17] Google: adapter.js; https://code.google.com/p/webrtc/source/browse/trunk/samples/js/base/adapter.js retrieved 20131215

[18] Google: Cube Slam; https://www.cubeslam.com/ retrieved 20131215

[19] Google: libjingle; https://developers.google.com/talk/libjingle/ retrieved 20131215

[20] Google: WebRTC Native APIs; February 2012, https://sites.google.com/site/webrtc/reference/native-apis retreived 20131127

[21] A. Grosskurth, M. W. Godfrey: A Reference Architecture for Web Browsers; Proceedings of the 21st IEEE International Conference on Software Maintenance (ICSM'05), September 2005, p661-664, http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=1510168

[22] C. Gutwin and M. Lippold, T. C. N. Graham: Real-Time Groupware in the Browser: Testing the Performance of Web-Based Networking; Proceedings of the ACM conference on Computer supported cooperative work (CSCW'11), March 2011, p167-176, http://dl.acm.org/citation.cfm?id=1958850

[23] M. Handley, V. Jacobson, C. Perkins: Session Description Protocol; IETF Network Working Group, July 2006, http://tools.ietf.org/html/rfc4566

[24] T. Hardie: Alternative Decision Making Processes for Consensus-Blocked Decisions in the IETF; IETF Network Working Group, October 2004, http://tools.ietf.org/html/rfc3929

[25] I. Hickson, D. Hyatt (eds.): HTML 5, A vocabulary and associated APIs for HTML and XHTML; W3C Working Draft, January 2008, http://www.w3.org/TR/2008/WD-html5-20080122/

[26] P. Hippensteel: Six things you need to know about WebRTC; Network World, July 2013, http://www.networkworld.com/news/2013/070113-webrtc-insider-271255.html

[27] IETF mailing list discussion: opportunity cost (was MTI video codec, charter, RFC 3929); November 2013, http://www.ietf.org/mail-archive/web/rtcweb/current/msg09713.html retrieved 20131127

[28] Influxis: XirSys; http://xirsys.com/ retrieved 20131215

[29] Instant IO Inc.: peerCDN; https://peercdn.com/ retrieved 20131215

[30] G. Kontaxis, D. Antoniades, I. Polakis, E. P. Markatos: An Empirical Study on the Security of Cross-Domain Policies in Rich Internet Applications; European Workshop on System Security (EUROSEC'11), April 2011, Paper No. 7, http://dl.acm.org/citation.cfm?id=1972551.1972558

[31] M. Kowalke: Don't Worry; Apple Will Soon Support WebRTC; WebRTC World, November 2013, http://www.webrtcworld.com/topics/webrtc-world/articles/360504-dont-worry-apple-will-soon-support-webrtc.htm retrieved 20131130

[32] T. Lammarsch, W. Aigner, A. Bertone, S. Miksch, T. Turic, J. Gaertner: A Comparison of Programming Platforms for Interactive Visualization in Web Browser Based Applications; 12th International Conference on Information Visualisation (IV'08), July 2008, http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=4577947

[33] T. Levent-Levi: How WebRTC Changes What RTC Really is; Author's personal blog, April 2013, http://bloggeek.me/webrtc-changes-rtc/ retrieved 20131130

[34] T. Levent-Levi: Let's Talk WebRTC Security; Author's personal blog, January 2013, http://bloggeek.me/webrtc-security/ retrieved 20131127

[35] R. Mahy, P. Matthews, J. Rosenberg: RFC 5766: State of Peer-to-Peer (P2P) Communication across Network Address Translators (NATs); Internet Engineering Task Force, April 2010, http://tools.ietf.org/html/rfc5766

[36] X. Marjou, S. Proust, K. Bogineni, R. Jesske, B. Feiten, L. Miao, E. Enrico, E. Berger: WebRTC audio codecs for interoperability with legacy networks draft-marjou-rtcweb-audio-codecs-for-interop-01; IETF Network Working Group, February 2013, http://tools.ietf.org/html/draft-marjou-rtcweb-audio-codecs-for-interop-01

[37] Mozilla: Media/WebRTC/Architecture; April 2013, https://wiki.mozilla.org/Media/WebRTC/Architecture retreived 20131127

[38] F. Mueller: Nokia files third patent infringement complaint targeting Google's VP8 video codec; FOSS Patents, May 2013, http://www.fosspatents.com/2013/05/nokia-files-third-patent-infringement.html retrieved 20131127

[39] Y. Namestnikov (eds.): IT Threat Evolution: Q3 2012; Kaspersky Lab ZAO, November 2012, http://www.securelist.com/en/analysis/204792250/IT_Threat_Evolution_Q3_2012

[40] B. Pedrick: What the CU-RTC-Web vs. WebRTC debate means for developers; TokBox Inc., http://www.tokbox.com/blog/what-the-cu-rtc-web-vs-webrtc-debate-means-for-developers/ retrieved 20131215

[41] Peer 5: Sharefest; https://www.sharefest.me/ retrieved 20131215

[42] H. C. Phuoc, R. Hunt, A. McKenzie: NAT Traversal Techniques in Peer-to-Peer Networks; In the proceedings of the New Zealand Computer Science Research Student Conference (NZCSRSC'08), April 2008, http://nzcsrsc08.canterbury.ac.nz/site/proceedings/Individual_Papers/pg242_NAT_Traversal_Techniques_in_Peer-to-Peer_Networks.pdf

[43] E. Popova: Vidtel Brings Video Closer to the Mass Market with WebRTC Support; Frost & Sullivan, January 2013, http://www.frost.com/c/10443/blog/blog-display.do?id=2324008 retrieved 20131130

[44] L. Quillio: Good News from Germany; The WebM Project, August 2013, http://blog.webmproject.org/2013/08/good-news-from-germany.html?m=0 retrieved 20131127

[45] J. Rosenberg: Bringing Real-Time Communications to the Web; April 2011, http://www.tmcnet.com/ucmag/features/articles/171983-bringing-real-time-communications-the-web.htm retreived 20131130

[46] J. Rosenberg: Cisco to open source its H.264 implementation and absorb MPEG-LA licensing fees; IETF mailing list announcement, October 2013, http://www.ietf.org/mail-archive/web/rtcweb/current/msg09269.html retrieved 20131127

[47] J. Rosenberg: RFC 5245: Interactive Connectivity Establishment (ICE): A Protocol for Network Address Translator (NAT) Traversal for Offer/Answer Protocols; Internet Engineering Task Force, April 2010, http://tools.ietf.org/html/rfc5245

[48] J. Rosenberg, B. Burman: H.264 as MTI for rtcweb; Internet Engineering Task Force, November 2013, http://tools.ietf.org/agenda/88/slides/slides-88-rtcweb-8.pdf

[49] J. Rosenberg, R. Mahy, P. Matthews, D. Wing: RFC 5389: Session Traversal Utilities for NAT (STUN); IETF Network Working Group, October 2008, http://tools.ietf.org/html/rfc5389

[50] P. Srisuresh, B. Ford, D. Kegel: State of Peer-to-Peer (P2P) Communication across Network Address Translators (NATs); IETF Network Working Group, March 2008, http://tools.ietf.org/html/rfc5128

[51] P. Srisuresh, M. Holdrege IP Network Address Translator (NAT) Terminology and Considerations; IETF Network Working Group, August 1999, http://tools.ietf.org/html/rfc2663

[52] Y. Takeda: Symmetric NAT Traversal using STUN; IETF, June 2003, http://tools.ietf.org/search/draft-takeda-symmetric-nat-traversal-00

[53] TenHands: TenHands; https://www.tenhands.net retrieved 20131215

[54] TokBox Inc.: TokBox; http://tokbox.com/ retrieved 20131215

[55] Y. Wei, D. Yamada, S. Yoshida, S. Goto: A New Method for Symmetric NAT Traversal in UDP and TCP; Asia Pacific Advanced Network (APAN'08) Network Research Workshop, 2008

[56] Song Zheng: Live video chat integration with FirePad.io, `http://songz.github.io/firepadvchat/` retrieved 20131215

[57] Zingaya Inc.: Zingaya, `http://zingaya.com/` retrieved 20131215

Chapter 4

Economic Aspects Of Network Neutrality

Sven Brunner, Sacha Uhlmann

Network neutrality is a very controversial subject that has been discussed extensively. The fate of the Internet is highly dependent on the outcome of the discussion and it might change the Internet as we know it permanently. This paper introduces the relevant stakeholders, discusses the current situation and the role of the regulator. It then continues to give an overview over four different scenarios: Strict network neutrality, termination fee scenario, user tiering and content and service provider tiering. It is shown that all scenarios lead to certain advantages and disadvantages for the stakeholders, and no perfect solution is in sight.

Contents

4.1 Introduction

In 1999 Cisco invented a router that was capable to differentiate network traffic, or more precise, the packages sent. A relatively easy way to perform "Deep Package Inspection" (DPI) was created. As a result, for network operators and broadband service providers it was now possible to prioritize, de-prioritize or even block packages [9]. This opened the way to actively manage Internet traffic. It is well known that the amount of traffic traveling through the Internet increases year by year and depending on the time of day the exiting capacity might not be sufficient anymore, resulting in delays or reduced quality of some services. Besides further increasing the capacity there was no mean to cope with this development. Since the innovation brought by this new technology, broadband service providers are enable to apply techniques to manage traffic. Because there are applications that are not depending on a low latency (for example e-mails) from a network operator's view it can make sense to give more sensitive application precedence (for example video telephony).
DPI opens a wide scope of additional possibilities. In cooperation with the network operators governments are now able to use lawful interception also with data traveling through the Internet [1]. The technique of managing and inspecting traffic is furthermore economically interesting. It enables to determine the source of a packet. Consequently, it is possible to assign a packet to a certain content provider, which enables network providers to treat packages differently depending on their source or content. This leads to potential economical and political issues, which are discussed in this paper.
First, the relevance of the subject will be motivated in Section 4.2. Then, in Section 4.3, the fundamentals of this topic will be introduced and an overview of several different definitions of network neutrality will be provided from several sources. Further the content of these definitions is discussed comprehensively and the differences between them highlighted. It is explained why it is discussed controversially and it is important for the Internet ecosystem. Another important aspect is to understand who the participants of the discussion are and their arguments. This is done in the analyzes of the most important stakeholders and their current views (Section 4.3.2). The description of the fundamentals will end in an explanation why the legal aspect and, therefore, also the regulator is inevitably intertwined with the discussion (Section 4.3.3).
The main part of this paper is the description and analysis of the current situation and four alternative scenarios, namely the strict network neutrality scenario (Section 4.4.2), the termination fee scenario (Section 4.4.3) and the Content and Service Providers (Section 4.4.4) and user tiering scenario (Section 4.4.5) respectively. The alternative scenarios are all analyzed with the same framework. It will described how these scenarios differ from the status quo and how the different stakeholders are influenced economically by them. Also the role of the regulator will be analyzed in terms of how it could help to implement or diminish drawbacks of the particular scenario. It is also discussed how the scenario could affect the overall welfare, the well-being of the society in economical sense, although in many cases no clear statement can be made. At the end, in Section 4.5, of this paper a summary of the observations and results and a brief outlook on how the discussion and the topic will evolve in the future will be given.

4.2 Motivation

The market has changed. While there where specialized network providers for different applications and services in the past, such as providers that only offered fixed telephone lines, cable TV or mobile connections. Many of these services were charged depending on how much the service was used, for example on a price per minute level. Since the early

Internet days, most of these providers also offer Internet connections. Today nearly all network providers offer flat rate based Internet access. This allowed other companies to offer their own services over the Internet, which are in direct competition to the services provided by the traditional network operators. Examples for such services include Netflix, Skype, and WhatsApp. The increased opportunities to deliver IP-based services to customers allowed network operators to offer a variety of services over their lines. Many earlier specialized operators offer IPTV, voice over IP (VoIP) and Internet over the same channel. Due to this increased competition network operators cannot achieve as much revenue as before.
The IP-based traffic has increased drastically. There is no end in sight for this enormous growth in volume accumulating. Cisco expects that the traffic volume in 2017 will be three times as much as in 2012, which is more than 1500 times the volume generated in 2000. Astonishingly, by far the largest proportion of this traffic is generated by private customers for video purposes [3, 4]. This phenomena is referred to as exaflood [14]. Due to this exaflood network operators must invest heavily into their infrastructure. Although the network equipment also increases in efficiency leading to lower prices for bandwidth, the increased demand still excels the gains in efficiency.
The increased demand in traffic and the broad availability of flat rates, as discussed above, lead to a situation, where network operators invest into an infrastructure, which mainly other market participants benefit from. Network providers try to counteract this situation by blocking and degrading traffic. 2005 the largest Swiss mobile provider "Swisscom" blocked the use of VoIP in their UMTS mobile network [21]. Other examples include Comcast, that degraded peer-to-peer (P2P) in their networks [14] and Verizon, that tried to charge extra fees for using tethering in their mobile network [22]. A research project by the Max Planck Institute for Software Systems called "Project Glasnost" suggested that 10% of all German Internet users experience P2P traffic degrading. It concluded that Kabel Deutschland, the largest network provider in Germany, are heavily using network management techniques. Project Glasnost has shown that nearly 40% of their traffic was scanned using DPI. In the United States even more traffic is inspected [14].
This behavior of the network operators started the discussion of network neutrality and the need for legal boundaries, and how network operators shall be allowed to manage their networks.

4.3 Fundamentals of Network Neutrality

Before the actual scenarios can be discussed, a common foundation must be built. Therefore, the following section will define the keywords, introduce the relevant stakeholders and explain why the legal perspective cannot be neglected. A stakeholder is a party that is influenced or can influence a project, business or technology.

4.3.1 Definition

Looking at the current discussion about the topic one can observe a broad range of meanings connected to the term "network neutrality". What does "neutral" actually mean? It is heavily dependent on who is defining the concept and what his or her interests are. In this Section several definitions are presented. The listing is in no way exhaustive and there exist many more definitions. It should only be shown that is not so clear what the term really means and, therefore, caution is needed when talking or writing about this topic. For that reason it will also be clearly explained which definition is used for the further discussion. The first definition is also the strongest form that is presented here and is

called strict network neutrality and often, but not exclusively, proposed from consumer rights groups [14]:

> "Net neutrality prohibits Internet service providers from speeding up, slowing down or blocking Internet traffic based on its source, ownership or destination."

What this definition basically says is that none of the technically feasible techniques of managing traffic may be applied by any Internet Service Providers (ISP). Packets must be transmitted according to the "Best Effort" principle [23]. This means that a packet arriving at a node should be forwarded as fast as possible in a first in - first out manner. This is actually how the underlying transport protocol, the IP protocol, was designed. When invented it was not meant to distinguish between packages of different applications and their potentially very different needs. It is obvious that this suits applications better which are not latency sensitive as sending e-mails or browsing the World Wide Web. [24]
A second and less strict definition is proposed from Robert Hahn and Scott Wallenstein [11]:

> "...broadband service providers charge consumers only once for Internet access, do not favor one content provider over another, and do not charge content providers for sending information over broadband lines to end users."

At a first glance the definition of strict network neutrality and this second definition seem very much alike. Service providers are also very limited in their possibilities of managing their network. However, there is one major difference. The definition by Hahn and Wallenstein allows the differentiation of packets from different applications. For example a provider could prioritize packets concerned with video streaming or Voice over IP (VoIP) in order to ensure a certain quality on the user side. This would be illegal according to the first definition. In the second one it is still not permitted to treat content differently regarding its source. Traffic from two different streaming services still must be handled equally.
The next definition is even less strict and is the first one proposed from a governmental agency, the Federal Communications Commission (FCC): [14]

> "A person engaged in the provision of fixed broadband Internet access service, insofar as such person is so engaged, shall[...]
>
> 1. Transparency "[...]publicly disclose accurate information regarding the network management practices, performance, and commercial terms[...]."(FCC, 2010, Sec. 54)
> 2. No Blocking "[...]not block lawful content, applications, services, or non-harmful devices, subject to reasonable network management."(FCC, 2010, Sec. 63)
> 3. No Unreasonable Discrimination "[...]not unreasonable discriminate in transmitting lawful network traffic over a consumer's broadband Internet access service."(FCC, 2010, Sec.68)[...]"

The FCC is "an independent US government agency" [7] and is responsible for regulating the communication markets (radio, television, wire, satellite and cable) [7]. Coming back to the definition, the first notable point is the highlighting of transparency. So broadband providers would not be as restricted as in the two previous cases but would have to clearly communicate how they manage traffic and which application or in general which traffic is favored. The third point the "no unreasonable discrimination" leaves room for interpretation. When is discrimination reasonable? Looking at the explanation from the FCC makes clear that this point is targeting illegal or harmful activities [14]. Thus, a

broadband operator can discriminate traffic to ensure the security of the network or services as parental control or just to prevent congestion of the network. However, there are still some unanswered questions. Is it allowed to degrade P2P traffic, which is sometimes, but by no means always, transmitting data that is violating copyright? Other question concerning degrading or blocking services, which could cause congestion problems could be raised. In connection an additional remark must be made. The definition explicitly excludes the wireless or mobile network services. The market for wireless network services in the United States is considered highly competitive and, therefore, does not need to be restricted [14]. What role competition plays and if or how it affects the topic of network neutrality as well as the consideration of the region specific differences will be discussed later in this paper.
We have seen several definitions of the term "network neutrality" and all define it differently. Now it is important to state which one is used when further referred to "network neutrality". We decided to go with the strictest definition, which was presented first. Of course, this does not mean that we consider this definition superior in any way or more accurate. One reason we chose this definition, is because it serves as model to one of our analyzes scenarios, the strict network neutrality scenario. Another factor is that this definition is desired from many strong network neutrality proponents and the most criticized and feared by the opponents.

4.3.2 Stakeholders

The market consists of several stakeholders. The first we will describe are the network operators. The network operators can be split into backbone and access operators [14]. Access Providers (IA) and there corresponding ISPs connect the users with the network. Most users are connected to one ISP and often do not have the possibility to switch this provider easily, either due to contractual constraints or due to little local competition [14]. Backbone operators are responsible for the interconnection between large network hubs and different access providers. Many large ISPs are their own backbone provider. However, for the network neutrality discussion mainly the last mile is relevant. Therefore, this paper will focus on the ISPs. As discussed in motivation, their point of view is that they are currently experiencing costs that mainly the other stakeholders profit from. They argue, that to guarantee a certain quality to the end users, they have to invest money, which they can currently not retrieve.
The next players in the debate are the Content and Service Providers (CSPs). This includes companies involved in producing content and services, as for example Google, Amazon, Yahoo etc. It can be argued that pure content providers, such as Hollywood studios, are a player on their own [9]. However, since most of them have their own service, we will not differentiate between pure Content Providers and CSPs. Same large CSPs often have their own backbone network infrastructure (e.g. Google). The CSPs do not see, why they should experience any additional costs [14]. They fear, that without any legal constraints, ISPs introduce managing techniques that are harmful for their business models. Many CSPs that offer products that are in direct competition with products of ISPs, such as online streaming platforms, fear that ISPs will gain an unfair competitive advantage by intentionally degrading their service.
Finally, there are the Internet users. It is assumed that currently over 2.7 billion people are connected to the Internet [5]. Of course, many Internet users also provide content to the Internet. For the usage of Internet users in this paper, we define Internet users as people who do not commercially produce content that is distributed the Internet. The Internet users only pay the ISP access fee to get access to all the CSPs. There is a minority of CSPs that also charge some fees directly to the Internet users for premium services, but these are not affected by the physical link between the two.

4.3.3 Legal Aspects and the Role of Regulation

While the focus of this paper is definitely not on the legal aspects, we believe that it cannot be ignored either. It is argued, that the first assessment of the network neutrality movement, which started the discussion was by Lawrence Lessig, a Stanford law professor [18]. The political discussion about network neutrality is highly controversial [14] and it will be the regulator's role, to ensure that if decisions are reached, they are implemented accordingly. For each scenario possible interventions by the regulator and their implementation are discussed. Even when a theoretical solution to the issues that arise with each scenario is found, the main challenge will be to find an efficient and robust legal implementation.

4.4 Frameworks and Scenarios

In the following section the current situation and possible scenarios are discussed. As Kraemer et. al. [14] suggests, the scenarios can be differentiated on which sides are charged and to which the network traffic is allowed to be influenced. The price regime can be divided into one-sided and two-sided pricing. In one-sided markets, the ISPs only charge the Internet users for Internet Access. Whereas in two-sided markets, the CSP is also charged for delivering the content over the last mile to the customer.
On the network regime side, it can be differentiated to what degree the ISPs can influence the the characteristics of the connection to the end user. Any new regulations shall focus on providing a stable environment for a competitive market.
Figure 4.1 shows the classification of the scenarios that will be discussed in this section according to the two dimensions

		Pricing regime	
		One-sided	**Two-sided**
Network regime	**Quality of service**	User tiering (IUs choose priority class.)	Content and service provider tiering (CSPs and/or IUs choose priority class.)
	Managed network	Status quo (Best effort network with traffic engineering and/or managed services.)	Termination fee (Additional fee for CSPs to terminate traffic at access ISP.)
	Capacity only	Strict net neutrality (No discrimination based on source, destination or content.)	

Figure 4.1: Overview of Scenarios [14]

4.4.1 Status Quo

As stated before traffic managing techniques are heavily used and implemented in today's network infrastructure. The ISPs argue that a smooth operation of the network would not be possible without them. This also means that traffic discrimination is a reality. Often it is used to ensure a certain level of quality of experience (QoE) for the customer [14]. Wu argues that there exist two kinds of discrimination that must be distinguished: Discrimination, which is justified, and discrimination that must be treated with caution [23]. In his opinion it should be allowed to discriminate traffic or applications that could harm the network (e.g. virus). However, when the discrimination is applied by the network operator to internalize cost, it would not be a legitimate argument. Economically speaking Wu argues that discrimination is reasonable when negative externalities can be avoided,

but certainly not when negative externalities are generated by the discrimination. The discrimination in both cases is not compatible with network neutrality.
A major concern of network neutrality proponents with discrimination is that some applications might be degraded [14]. As seen in the motivation (Section 4.2) this is done quite extensively, often in connection with P2P traffic, but not exclusively. It is feared that an ISP could degrade a certain application in order to favor an own service, as presented in the examples in the motivation (Section 4.2). The second concern is related to blocking, meaning that ISPs hinder certain information to be delivered to the end users. This could potentially also limit freedom of speech [14]. Again this could also be used to exclude rival applications form the own network or force CSPs to pay fee in order to stay reachable. It is questionable that to interfere in such a strong way is in the interest of an ISP. A damaging loss of reputation would most certainly be the consequence [14]. Ed Whitacre, chairman of AT&T, is convinced that if his company started to block applications or information, it would drive its customers to competitors immediately [9]. This implies that there is enough competition in the market and, therefore, a company engaging in blocking or unreasonable discrimination would be punished by the customers. The existence, however, of this competition is doubted. Friedman argues that only big content providers are able to choose between different ISPs, because they are mostly connected directly to the Internet backbone. Friedman claims the end users might not be in such a comfortable position [8]. Hahn and Wallenstein however stated, based on a statistics from the FCC, that in June 2005 almost 90 percent of American households had the choice between two or more ISPs. But they also recognize that this does not necessarily mean that these ISPs are in direct competition [11]. So whether or not there is enough competition in United States' fixed wire market remains somewhat unclear. Of course, in other countries the situation might look different. Even if there was enough competition, it might still not prevent ISPs to examine degrading or blocking techniques. The US mobile phone market is considered highly competitive [14]. Despite this fact, Wu found many examples of "non neutral behavior" [14] in his analysis of broadband usage restrictions [23].
In connection with competition another aspect must also be considered: Transparency. In this context a transparent behavior would imply to disclose information about traffic management openly. This can be done with two different approaches: Top-down or bottom-up [14]. In the top-down approach the ISP itself is responsible to provide this information clearly to its customers and the public. Also it is important to deliver this information in understandably manner for the consumers. With the second approach, bottom-up, the responsibility lies with the customer. The ISP just needs them to be able to examine whether traffic management techniques are applied and if yes which ones. This is exactly what the "Glasnost Project", presented in the motivation (Section 4.2), does. Generally it can be said that the ISPs have much room for improvement concerning the aspect of transparency. How does transparency relate to competition stated earlier? Let's imagine a customer is provided with all the information he or she wants about his or her ISPs traffic management and does not agree with it. If there is no other ISP to switch to, the customer does not benefit from the transparency.
The next important aspect to consider in today's situation is innovation. Generally, one can distinguish between the innovation at the core, which means the network infrastructure, and the innovation at the edge, the applications and services [17]. Here the commonalities in opinion and arguments end. Some, like van Schewick [17], argue that innovation at the edge is more important than at the core [17]. Others, as Sidak and Teece [18], on the contrary conclude that innovation on the application side is only possible if there is technical progress at the core [18]. The former argumentation supports the idea that the market should be regulated in order to ensure network neutrality, the latter the ISPs to desire develop additional revenue. Why do these ideas about innovation differ? The Internet is not designed to support a specific application. It is sometimes even con-

sidered ”dumb” in a sense that it does not care what traffic travels through its nodes, this design is often called ”end-to-end” design or principle [9]. The intelligence comes with the applications at the end of the network. For many network neutrality advocates and network experts it is indisputable that the Internet had never become such an important part of modern society and could not have grown so enormously if it were designed differently and, therefore, the architecture must remain the same. They see the advantage in current architecture in the fact that every application is treated equally. Whereas the story of the Internet is undoubtedly a successful one there are also voices who believe that the principles should be adapted to today's changed economical and technical environment. Hahn and Wallenstein for example think that applications could become even more innovative if they had to ”respond to price signals from platform providers” [11]. They also mention that it is just not known whether another approach would be better, because it was never tried. Sidak and Teece believe, that applications that rely on low latency or jitter would never be developed, if the ISPs were not allowed to offer QoS transactions [18]. They further ensure that QoS produces more heterogeneous content and applications of which also customers benefit. In a heterogeneous market the competitors can differentiate their services and products, which leads to more intense competition and lower entry barriers. Low entry barriers and possibility to differentiate also help to reduce the danger of a monopolistic market. [18]
So far, no global QoS standards have been defined. Meanwhile, the CSPs are already seeking possibilities to enhance the QoE for customers [14]. A common solution to achieve this is work with so-called content distribution networks (CDNs). These are often independent companies possessing their own infrastructure. This infrastructure is built supplemental to the already existing one owned by the ISPs. CSPs may then send their traffic through a CDNs network to avoid routes prone to congestion [14]. Some very big CSPs even have their own capabilities to provide the very same feature [18]. While QoS clearly violates network neutrality, with CDN's it is not so obvious. Despite their important role in the status quo, they have almost entirely been excluded from the discussion about network neutrality [18]. Actually, the mechanisms and results when applying QoS or when working with CDNs are similar and it seems questionable whether they should be differentiated.
As mentioned in the motivation (Section 4.2), ISPs are trying to find additional revenue streams. We will now have a quick look at the current revenue streams, as illustrated in Figure 4.2, in order to be able to compare the today's situation with the alternative scenarios.

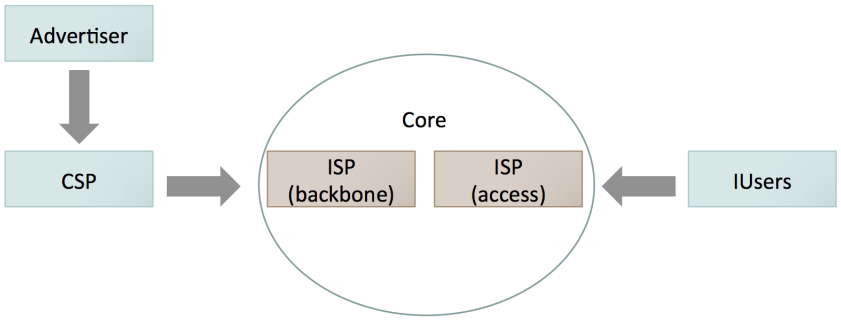

Figure 4.2: Current Revenue Streams [14]

There are three major streams relevant to the topic. The first stream concerns the CSPs and the advertisers. CSPs sell advertising space to companies and the price advertisers are willing to pay is mainly depending on the how many users are attracted by a certain CSP. Obviously the service or content must be available and, therefore, the CSP must pay one or several (backbone) ISPs to be interconnected with the Internet. This represents the second stream. The third revenue stream is the one probably best known. The end

users paying for access to the Internet. They are charged by the access ISPs. The amount of the access fee usually depends on the bandwidth the user orders. Of course, there are other income possibilities for the different stakeholders. Some CSPs for example charge customers for using theirs services or consuming their content. However, for the further discussion only the three described streams are essential. [14]
It is quite safe to say that there are going to be major changes in the near future. But the question remains: Into what direction will the Internet advance? We will not be able to give a concluding answer to this question but we will examine how a potential future could look like. Therefore, we analyze four alternative scenarios proposed from Kraemer, Wiewiorra and Weinhardt [14].

4.4.2 Strict Network Neutrality

In the following section the strict network neutrality is introduced. After a description, the economic influence on the different stakeholders, the role of the regulator, and the impact on the welfare are discussed.

4.4.2.1 Description

The strict network scenario is the implementation of the strictest definition. Therefore, ISPs are not allowed to speed up, slow down or block traffic in any form (Section 4.3.1). Since absolutely no network management by the ISPs would be allowed any longer, it is argued that it could lead to congestion problems during peak times. Furthermore, ISPs can no longer guarantee a certain quality for alternative services such as IPTV [14].
As shown in Figure 4.3, the revenue streams in the strict network neutrality scenario are the same as in the current situation.

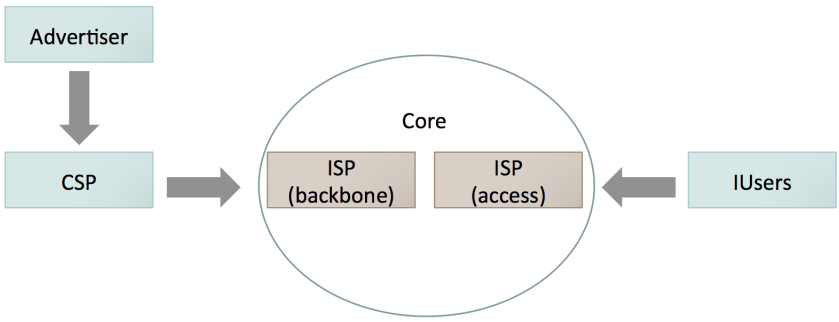

Figure 4.3: Revenue Streams in Strict Network Neutrality Scenario [14]

4.4.2.2 Economic Influence on Stakeholders

The strict network neutrality scenario opposes many issues for ISPs. Since they cannot guarantee a quality for alternative services, it is questionable if they can offer them at all. If they cannot, revenues for the ISPs will be reduced. In order to increase revenue, needed to invest into their network infrastructure, ISPs will have to increase their prices to the end customers. Unlike in the non-network neutrality scenarios, they cannot get any additional revenue streams from any other sources. The competition between different ISPs is also strongly limited, as they can only differentiate volume and bandwidth in their offers. Only for very few ISPs the strict network neutrality scenario can actually be an opportunity: It is argued that for some high quality ISPs it could result in reduced competition [16]. The lower quality ISPs could not compete at all in such a market, as they could not compensate this disadvantage by enhancing quality of transportation [14]. For many CSPs, the strict network neutrality scenario has a positive effect. They will have a legally guaranteed best effort connection to every Internet user. They furthermore do not

have to fear any extra charges by the ISPs for the delivery of their content to the end users. The threats for the CSPs lie in the connection quality. Since the ISPs might not have the assets to ensure a certain quality in their networks, congestion issues may arise. This can lead to a lower quality of experience for the end users. The CSPs now have the issue that they cannot deliver their content in the expected quality, which consequently would lead to fewer customers. Finally, the incentive for CSPs to develop new real time application is not given any longer. For the Internet user the strict network neutrality scenario has the advantage that no extra charges by the ISP have to be feared. As every package is treated equally, it does not matter what Internet is used for. There is no degraded quality for certain services or application and especially no blocked services. However, this also means that real time applications cannot be treated differently. The experience of IPTV, VoIP, Online Gaming etc. might drastically be reduced or not possible in an acceptable quality during peak hours. The fact that ISPs cannot seek any other revenue streams might directly lead to an overall price increase for Internet access. Internet users with reduced needs would not find any offers that are suitable offer for them. And probably worst, a strict network neutrality approach might lead to a possible reduction in new innovative real time applications, as the incentive to produce such would be reduced.

4.4.2.3 Role of the Regulator

As this scenario is the strictest, it is up to the regulator to change the regulations and implement them. As discussed before, this scenario could hurt the market as it is today. Therefore, it is likely that regulators would not follow this very strict approach, but find a reasonable compromise to enforce network neutrality.

4.4.2.4 Impact on Welfare

The effect on the overall welfare in a strict network neutrality scenario is hard to assess. The positive effects include that it ensures freedom of speech and prohibits ISPs from gaining unfair competitive advantage by blocking or degrading services and content of competitors. However, as we have seen a direct implementation of the strict network neutrality scenario leads to many negative effects on all stakeholders. This includes reduced income revenue streams for ISPs, decreased customer numbers for CSPs and reduced quality in time sensitive application for customers.

4.4.2.5 Critical Claim

While the very strict scenario definitely has its advantages, it just does not seem very realistic to us. The predicted raise in bandwidth foreseen for the next years, makes investments by ISPs necessary. With strict network neutrality, the incentives for ISPs is too low to invest enough to ensure a certain quality. It is questionable, if the discussion changes the situation in countries where it governments actually suppress freedom of speech. Especially in Europe the past has shown that if ISPs try any unfair degradation of services, blocking certain apps etc. the outcry by the customer base has been loud enough so that the market regulates itself.

4.4.3 Termination Fee

This section analyzes the termination fee scenario. The termination fee scenario is presented with its economic consequences on the relevant stakeholders, the general welfare, and the regulatory role. Unlike in the strict network neutrality scenario, the termination fee scenario would allow for additional revenue streams for the ISPs.

4.4.3.1 Description

The termination fee scenario is one of two scenarios, where the ISPs act as a two-market operator. Analog to a credit card provider, the ISP on one side charges the Internet user for accessing the web over their network and, on the other side, the ISPs also charge the CSPs for delivering their content to the end users. Unlike in other non-network neutrality scenarios, this additional financial burden does not result in any immediate reward for the CSP. They are charged independently of how their traffic is managed by the ISP [14]. In terms of revenue streams, visualized in Figure 4.4, this means that the network providers get additional revenue by the CSPs. Kraemer et. al. argue, that the risks of status quo still apply to any non-network neutrality scenarios [14]. ISPs might still want to block or degrade the quality of the services of competitors. While it can be argued, that the incentive to degrade traffic of competitors is lowered, since they have additional revenues, this can hardly be argued for blocking any sort of freedom of speech. If any provider has an incentive to block a website, this will not change under any non-network neutrality scenario. [14]

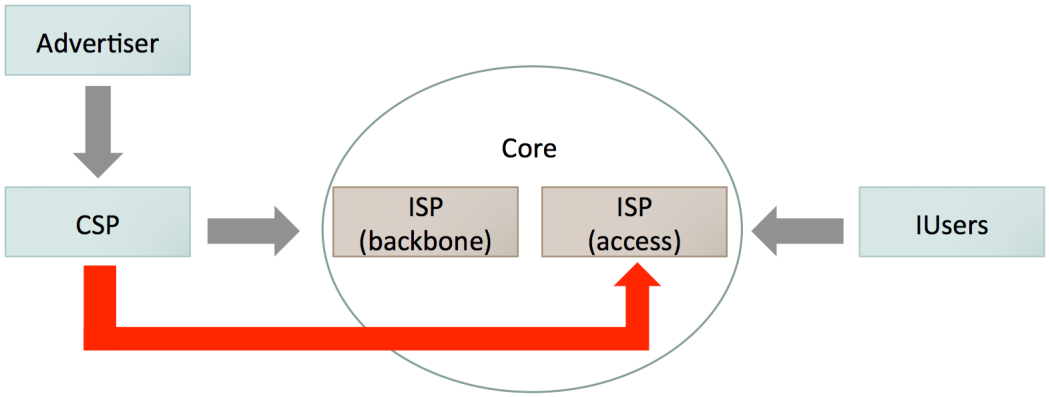

Figure 4.4: Revenue Streams in Termination Fee Scenario. Based on [14]

4.4.3.2 Economic Influence on Stakeholders

For ISPs the termination fee scenario allows them to have an additional revenue stream without delivering them anything new, so implementation for the ISPs should be rather easy. However, these extra charges for the CSP might decrease the number of active CSPs in the market. Over-charging them can lead to an "Tragedy of Commons" situation, where all ISPs try to charge the CSPs as much as they can. The resulting financial burden might push the CSPs off the market, resulting in less profit for all ISP. The CSPs are threatened by additional fees, without generating any immediate benefit to the situation today. This is especially dangerous for CSPs that are already struggling today to get enough revenue to succeed in the market and start-ups. Internet start-ups are often already struggling funding their operations, will need even more money before generating income. On the other hand it can be argued, that the incentive for the ISP to block or degrade CSP traffic is decreased. This leads to a lower risk that ISPs experience such behavior. It can also be argued, that the ISPs would use their additional revenue to invest into the network structure. Therefore, the CSPs would have a better connection to the Internet user, leading to higher customer satisfaction. It is also likely, that since the cost are split between Internet users and the CSPs, that ISPs charge lower access fees to the user and attract more customers. The advantage for the customer is clearly, that it is unlikely that they will be charged higher prices. The threats are closely related to the threats of the CSPs, i.e. that if CSP leave the market or new CSP do not enter the market due to

increased cost, and Internet users will therefore experience a reduction in content variety and less new services. It is also possible that CSPs will seek additional revenue streams, and, therefore, more CSPs try to sell "premium services", where the end users are directly charged, or that they increase the amount of advertisement in their content/services.

4.4.3.3 Role of the Regulator

As in the strict network scenario, the regulators' role is to provide a regulatory framework. The risks described in 4.4.3.2 must be considered specially by the regulator, making sure that pricing is fair and that taking care of any antitrust issues are taken care of. Finally, the regulator must consider implementing special laws for new CSPs to ensure that the scenario does not become an innovation blocker.

4.4.3.4 Impact on Welfare

While it is often argued that any non-network neutrality model with two-sided pricing will result in a decreased number of content and services provided by the CSPs [14], resulting in less innovation in the Internet and decreased welfare, there is a growing amount of scientific research showing that the opposite could be the case. Economides and Tag show that Internet users and ISPs are better off with non-network neutrality [6]. Their model is based on the assumption that an additional Internet users is worth more to a CSP than an additional CSP to an Internet user. With a termination fee, pricing for the Internet user can be decreased, attracting more Internet users, which then again leads to more customers for the CSP. Hahn and Wallenstein argue that in a terminating fee scenario, the waterbed effect would lead to rebalancing of fees [10]. The waterbed can be explained as that when one side of the market is charged less, the other charged more.

4.4.3.5 Critical Claim

The termination fee approach allows for a fairer distribution of costs between the different stakeholders. However, pushing costs towards the CSPs without giving them any extra services results in tremendous opposition. It would also mean that ISPs and CSPs have to operate with contracts. It is unrealistic that this is possible for especially for smaller ISPs and CSPs, unless implemented by global agreements.

4.4.4 CSP Tiering

In this section the CSP tiering scenario is first introduced and then portrayed under various aspects. In contrast to the previous scenarios, the CSP tiering scenario allows differentiation at the consumer side of the market.

4.4.4.1 Description

As we have seen the Internet offers space for a large number of diverse and sophisticated applications. As a result, the requirements for network quality differ heavily between the applications [14]. One can roughly divide them into two groups: Applications that severely suffer from congestion (congestion-sensitive) in terms of reduced quality and applications that are not (congestion-insensitive) [14]. The services and applications not affected are also called "elastic services" [15], because lost packets can be "reordered from the source" [15], examples are email or browsing the web. On the other hand, we have the congestion-sensitive applications, that in case of congestion are likely to have reduced quality or fail. Members of this group are all interactive services as VoIP, online gaming or video streaming [15]. So for providers of such services there is certainly a need to be

able to guarantee a certain quality. In a best effort network, where every application is treated equally, this would not be possible. Consequently, prioritizing certain traffic requires another model than the strict network neutrality scenario.
Such a model is referred to as CSP tiering [14]. In this model or scenario CSPs are allowed to pay ISPs for priority access. This simply means to introduce true QoS. Of course this is not conform to strict network neutrality. There are several possibilities of how such a model could be implemented. As proposed by Hermalin and Katz, an ISP could offer several different levels of quality. The higher the quality a CSP purchases the higher the fee [12]. CSPs whose content is valued higher by the customers are eager to buy a quality as high as possible. [19]
The model considered to be the most relevant is the one in which ISPs offer two lanes with different quality [14]. One lane is working according to the best effort principle and remains free of charge. The second lane is the premium lane where packets are prioritized and CSP need to pay to have their data sent in the lane. Although it is helpful to imagine these two different approaches as two separate lanes, it is needless to say that for the technical implementation this would not mean to physically build two disjoint networks. If there is data arriving at a node coming from a service that paid for priority, other packets simply would have to wait, all applications still share the same network. What impact does this have on the average waiting time of a packet? With help of queuing theory it is relatively easy to show that priority packets have to wait less, in contrast other packets longer compared to the situation where all packets are treated equally [14]. Of course this supposes that not all CSPs buy the priority access.
This leads us to one of the major concerns regarding CSP tiering. Because the best effort lane does not generate revenue for the ISPs they might have the incentive to degrade the quality. If this degradation is intense enough all CSPs may be forced to buy the premium lane. This is commonly referred to as dirt road fallacy [18] which will be discussed in greater detail in the next Section.
With this scenario the same new revenue stream for the ISPs is generated as already described under the termination fee scenario, see Figure 4.5. However, there is a major difference between the two scenarios. Whereas in the termination fee model the CSPs receive nothing in return for paying the ISPs in the CSP tiering scenario they are provided with a higher transmission quality for their services.

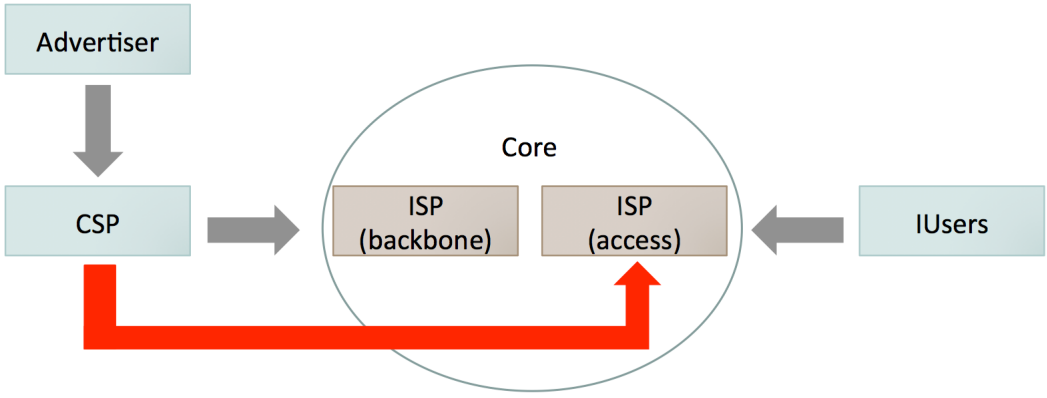

Figure 4.5: Revenue Streams in CSP Tiering Scenario. Based on [14]

4.4.4.2 Economic Influence on Stakeholders

In the CSP tiering model ISPs have similar opportunities as in the termination fee scenario. They unquestionably generate more revenue under the assumption that there are some CSPs that pay for the premium lane. This additional revenue then could be used to invest

further into the network which has a positive effect on the network's quality possibly leading to even higher profits (because CSPs are ready to pay more for a better network) [19]. A potential problem is the so called re-congestion effect [14]. This means that CSPs that have priority access are valued higher by customers than before and, therefore, generate also more traffic. This increased traffic again leads to congestion or in this context re-congestion. As many opportunities and threats presented in this paper, this is mainly an assumption and cannot be proven until such a QoS model were implemented in reality.
Like the name of the scenario suggests the CSPs are confronted with most changes. A potential threat comes with the fees the CSPs have to pay for faster transmission. Some fear, a prominent exponent is Lawrence Lessing, professor for law at Stanford University, that this would kill innovation at the application layer. He argues that especially for start-ups, the future Facebooks and Googles, it is almost impossible to compete with already established companies when they have to pay in order to reach their users. [9]
As mentioned before, the biggest threat to be considered is the dirt road fallacy where all CSPs are pushed into the priority lane due to poor quality of the best effort lane. The result would just be a termination fee model. For many network neutrality proponents this is very likely to happen. Sidak and Teece, however, do not see any reason to be afraid. They argue that the market for Internet content is global and the ISP are at most controlling a certain local region. Which means that a single ISP could never reduce the quality of certain service as a whole and, therefore, does not have the incentive to do so in the first place. They go even further by saying that reduced quality of a service does not lead to significant loss of customers which stands in strong contrast to many experts' opinions. [18]
Others admit that the dirt road fallacy really is a potential issue, but see a remedy in a minimum quality standard established by the regulator [14]. This argument will be discussed in the next Section. Besides the threats there are also some opportunities for the CSPs. Some see quality of the content delivered by a CSP and the quality of transportation either as complements or as direct substitutes [14]. Assuming they are substitutes this is a chance by lower valued CSPs. "Lower value" meaning that the users do value the level of service as high as the content from other CSPs. An example could be a smaller choice of available movies from a streaming service compared with its competitors. Such a CSP could gain in attractiveness when buying priority access.A further positive aspect is the ability for a CSP to further differentiate its products through optional QoS. As already mentioned before, many see in allowing QoS the opportunity for brand-new applications and services otherwise not realizable. [18]
The users would probably not be affected too strongly. The biggest benefit for them is definitely the enhanced quality of some services from CSPs that purchased the priority lane. When following the argumentation that CSP tiering favors big CSPs the content variety for users could be reduced. On the other hand, application may be available that otherwise would not have been. Some fear that the CSPs could pass on the additional cost under this regime to the users in form of higher fees or even charge users for content that has been free so far. [15]

4.4.4.3 Role of the Regulator

In this scenario the regulator most likely would be responsible for avoiding the dirt road fallacy. A large majority sees the solution in establishing a minimum quality standard (MQS). This means that the best effort lane still must deliver data in acceptable quality. The problem here is how much quality is enough. Established ISPs might be interested in setting this standard very high in order to use this as an entry barrier for new market entrants.

According to another argumentation the ISPs would then have to invest more into the network as would be optimal which results in a welfare loss [14]. For a regulator it would also be quite difficult to define an appropriate quality standard which is neither too high nor too low. Another question is how to describe such a standard: Mbit per second? Duration of loading a web page? Additionally such a standard is only reasonable if it can be enforced and violations against it can be punished. Despite the problem that could arise many experts still think this is manageable and the only realistic solution to reduce the drawbacks of this scenario. [14]

4.4.4.4 Impact on Welfare

What effect the CSP tiering scenario would have on the overall welfare is heavily depending on assumptions. Would ISPs reinvest their additional revenue in the network? Would the competition among CSPs increase or are the large companies favored? Most important: Could the dirt road fallacy be avoided? No one is able to give a final answer. Nevertheless, most experts think that CSP tiering is most likely to be welfare enhancing or does less harm than strict network neutrality. [14]

4.4.4.5 Critical Claim

In our opinion this model is superior to the previous two. Where in the termination fee model simply a fee is established and with strict network neutrality ISPs must give their private propriety away for free with CSP tiering there is really a chance for all participants. ISPs can generate more profits, CSPs have the possibility to guarantee a certain quality to their customers and the users benefit form richer and more applications. We also think that MQS would be a nice solution for the main threats but this needs to be done carefully and in consultation with all the relevant stakeholders.

4.4.5 User Tiering

In the following section the last scenario, the user tiering scenario, is examined regarding its potential concerns and opportunities. Like in the CSP tiering scenario, the ISPs can also differentiate the prices at the consumer side of the market.

4.4.5.1 Description

The user tiering scenario is the counterpart of the CSP tiering on the user side. Here CSPs do not pay any additional fees contrary to the users. The revenue stream are visualize in 4.6. In this scenario again many models are imaginable. An interesting idea is to split the available bandwidth into lanes similar to what we have seen in the CSP tiering scenario, this with the background of the Paris Metro pricing [2]. There existed two classes in the Paris Metro, one twice as expensive as the other one. So far this is a well-known principle used by all sorts of means of transport operators. Where normally a passenger in the first class is provided with a benefit (for example better service or more comfortable seats), in the Paris Metro the cars of the two classes did not differ at all. However, due to the higher price there were less people in the first class and, therefore, also more room. This would be exactly the same principle as having two bandwidth lanes, just that one is more expensive and as a result less traffic would occur which would lead to reduced danger of congestion. However, the approach in which users can choose applications and services to be prioritized seems to be the most probable [14]. Of course, the user has to pay an additional fee to the ISP. In this scenario many fears and concerns of a non neutral network will not occur. Because CSPs are not charged, no negative effects on innovation will occur. Some think that user tiering would not even be violating network neutrality

because it does not change the ”one-sided pricing paradigm of the Internet” [14], but if strict network neutrality was implemented user tiering would clearly be illegal.

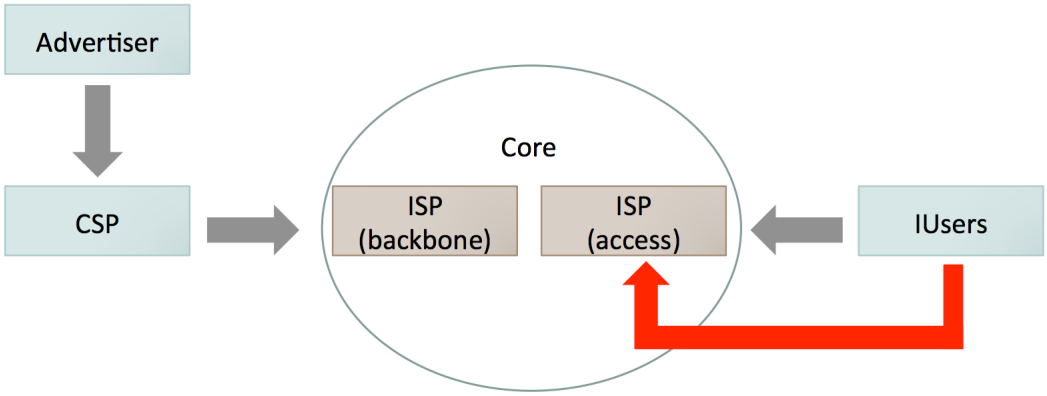

Figure 4.6: Revenue Streams in User Tiering Scenario. Based on [14]

4.4.5.2 Economic Influence on Stakeholders

The ISPs have the same opportunities as in the user tiering scenario, because they have the same options, but are charging the other side of the market. This leads again to higher investment possibilities. The biggest threat in this scenario for the ISP are negative user reactions. This fear is very real, and many ISPs actually prefer the CSP tiering. The relation between the Internet users and ISPs is more on a psychological and emotional level, than the rational relation level between the CSPs and ISPs. They fear that introducing new charges to the Internet users will damage this relation a lot more than a CSP tiering would and result in a huge image loss [14]. This image loss might lead to a customer changing their ISP, even though it is not based on rational reasoning.
CSPs would not experience any extra charges and, therefore, the risk of less innovation in the Internet is also not given. Start-ups do not have to seek additional financial assets at the beginning. Under the assumption that the services which are not prioritized by the users, will decrease in quality over the next years (for example due to the exaflood), it is possible that the CSPs lose current customers. There might be Internet users, that use or would like to use the service of the CSP, but their benefit from using it, is lower than compared to the price for the prioritized access at the ISP. If their benefit is smaller compared to the price they would have to pay and the quality of experience suffering they might not use the service at all any longer. [14]
Of course Internet users would be confronted with additional costs. An even more realistic threat is, though, that normal customers, who are not willing to play premium prices for certain products, would be neglected by the ISPs as their focus would be on premium Internet users. They might bundle their premium lane with other products, resulting in decreased offerings for users with specific needs. Furthermore, the ISPs could increase the premium bandwidth fraction of their physical lanes over the years, resulting in worse quality for the best effort lane. Positive for the Internet users is, that they can choose which services are valuable to them and it is quite realistic, that the ISPs then really would provide them with the expected quality. Also, since this scenario has little effect on the CSPs, the risk of loosing content and service providers, including new innovative ones, is not given [14]. Users that do wish any premium quality might find offers by ISPs that are better suited to them, i.e. cheaper than the current options. [23]

4.4.5.3 Role of the Regulator

As in the CSP tiering scenario, the regulators' role is to set a certain MQS. The only difference is, that this time it shall ensure a certain quality for non premium Internet users and not for CSPs.

4.4.5.4 Impact on Welfare

The effect on welfare is hard to estimate, especially since the reactions of the Internet users are difficult to assess. It has been shown by Schwartz, Shetty and Chiu that two service classes can be socially beneficial [20]. However, it is unclear to what extent the same effect will apply in the user tiering scenario. If a user tiering scenario is considered, a MQS must be set to increase the likelihood of a positive welfare effect.

4.4.5.5 Critical Claim

In our opinion this this scenario is not to be considered realistic at the moment. The reaction of the Internet users in the last years on any trial to devalue their situation led to a huge outcry, and the ISPs withdraw their proposals. The ISPs fear loss of reputation too much to introduce costs, which cannot be understood by the users at the moment. One author is of the opinion that, that if in the future the quality of experience in the Internet declines substantially, the perception of the Internet users would change rapidly, allowing ISPs to introduce such a fast lane without any major opposition.

4.5 Outlook and Conclusion

By now only two countries worldwide have established a law concerning network neutrality. In 2010 Chile adopted a law which was implemented the preceding year. An early version of this law planned to prohibit any form of traffic management or QoS or in other terms it would have been the very first implementation of network neutrality in its strictest form. In the final version though interests of both sides were taken into account and a tiering system was not declared illegal. In 2011 Dutch mobile providers expressed the intention to raise additional fees for using certain messaging and VoIP services (for instance WhatsApp or Skype). As a direct result a law was adopted to prevent the providers from doing so. In addition, it is interesting to see that the law explicitly regulates the mobile market whereas in the USA the discussion is mainly concentrated around the fixed network market. [14]
A few month ago the German government promised to invest heavily in the network infrastructure in the entire country. It is planned to provide up to a billion euros per year. The government also stated that a goal of the next legislation would be to adopt network neutrality as a law. [13]
The European Union developed similar rules to the ones stated by the FCC which served us as one potential definition of network neutrality. In the European framework a tiered system is not prohibited per se. However, the framework proposes minimum quality standard as precautionary measure. [14]
As mentioned before there are huge differences between different countries in terms of how much the topic of network neutrality is discussed. The discussion is mainly located in Europe and the United States, especially in the US since most big CSPs are located there. As discussed, the CSPs are the main opponents of non-network neutrality scenarios as they try to defend their current business models. [9]
It is likely, that the discussion will spread in the future to other topics connected to the Internet ecosystem, especially towards device neutrality and content neutrality [13]. Device neutrality is about how much control mobile operating systems developers can

exercise over what can be installed on their devices through their app stores, how own programs are pre-bundled and how open the devices are to technologies of other vendors. Content neutrality is not about devices, but how content can be found over the Internet. Due to the market power that, for example Google has in the search engine market, it is argued that all search results must be based on the same fair principles [14]. As we have seen, the subject of network neutrality is complex and it is unlikely that the interest of all different stakeholders can be reconciled. Traffic management by ISPs are reality these days, often not even communicated transparently to the Internet users. Internet users and especially CSPs fear that traffic management will lead to disadvantages and, therefore, try to enforce a explicit legal framework. In the view of the ISPs the current situation is not future proof and, they would need additional revenue sources to ensure a certain quality for their customers. Especially, since an increase in traffic volume is expected in the next years. The ISPs argue, that primeraly CSPs profit from their investments into the infrastructure without paying their share.
Several alternative scenarios are introduced and discussed, but there is no solution which satisfies all needs. An isolated view on just one or two scenarios would not embrace the variety of stakeholders and their divergent interest. The strict network neutrality scenario (Section 4.4.2) leads to many issues on all sides and its implementation can be doubted, as it requires very special terms. In comparison to all other scenarios, it prohibits any network management techniques, i.e. it is even stricter than the status quo. Any charges except for access are not allowed. The user tiering scenario (Section 4.4.5), is rather unlikely to happen in the current situation, due to the expected opposition of the users. In general it is comparable to the CSP tiering scenario, with similar legal issues and solutions. The only difference is which side is charged (users or CSPs). As users are more emotional, the expected opposition would be higher. The termination fee model (Section 4.4.3) does not offer any direct benefit to the CSPs, and it is argued that it could result in less innovation in the Internet. In the CSP tiering scenario (Section 4.4.4) on the other hand, the CSPs get additional value for the extra fees in return, which is the main difference between the two scenarios.
To conclude, neither the current situation nor any scenario are perfect. The discussion is far from over and it is unlikely, that any scenario is implemented in a pure form. Any legally binding implementation will most probably be based on a compromise of the different scenarios.

Bibliography

[1] N. Anderson: *Deep Packet Inspection meets 'Net neutrality, CALEA*, arstechnica, 25.7.2007. Last retrieved 19.11.2013 from http://arstechnica.com/gadgets/2007/07/deep-packet-inspection-meets-net-neutrality/.

[2] C.-K. Chau, Q. Wang, D.-M. Chiu: *On the Viability of Paris Metro Pricing for Communication and Service Networks*, in Proceedings of the INFOCOM Conference, March 2010.

[3] CISCO: *Cisco Visual Networking Index: Global Mobile Data. Traffic Forecast Update, 2012-2017*, February 2013. Last retrieved 10.11.2013 from http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537 /ns705/ns827/white_paper_c11-481360.pdf.

[4] CISCO: *The Zettabyte Era - Trends and Analysis*, Mai 2013. Last retrieved 10.11.2013 from http://www.cisco.com/en/US/solutions/collateral/ns341 /ns525/ns537/ns705/ns827/VNI_Hyperconnectivity_WP.pdf

[5] ICT: *ICT Facts & Figures 2013*. Last retrieved 10.11.2013 from http://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2013.pdf

[6] N. Economides, J. Tag: *Network neutrality on the Internet: A two-sided market analyses*, Information Economics and Policy, January 2012.

[7] Federal Communications Commission. *FCC - What we do*. Last retrieved 11.11.2013 from http://www.fcc.gov/what-we-do

[8] R. Friedman: *A Primer on network neutrality*, Intereconomics, January/February 2008.

[9] P. Ganley and B. Allgrove: *Net Neutrality: A User's Guide*, Computer Law & Security Review, Volume 22, Issue 6, pp. 454-463, August 2006.

[10] C. Genakos, T. Valletti: *Regulating prices in two-sided markets: The waterbed experience in mobile telephony*, Telecommunications Policy, February 2012.

[11] R. Hahn and S. Wallsten: *The Economics of Net Neutrality*, The Economists' Voice, No. 3, pp. 1 - 7, June 2007.

[12] B. E. Hermalin, M. L. Katz: *The Economics of Product-Line Restrictions with an Application to the network neutrality Debate*, Informationen Economics and Policy Vol. 19, pp. 215-248, February 2007.

[13] D. Kuhr, G. Bohsem, M. Bauchmueller, T. Oechsner: *Schwarz-Rot will Breitband-Internet und Forschung foerdern*. Sueddeutsche, Nov 2013. Last retrieved 13.11.2013 http://www.sueddeutsche.de/politik/koalitionsverhandlungen-schwarz-rot-will-breitband-Internet-und-forschung-foerdern-1.1810534

[14] J. Krämer, L. Wiewiorra, Christof Weinhardt: *Net Neutrality: A Progress Report*, Telecommunications Policy, Telecommunications Policy 2012, November, 2012.

[15] J. Kruse: *Network neutrality and quality of service*, Intereconomics - Review of European Economic Policy, Vol. 43, pp. 25-30, January/February 2008.

[16] P. Njoroge et al.: *Investment in Two-Sided Markets and the Net Neutrality Debate*, May 2012.

[17] B. van Schewick: *Towards an Economic Framework for network neutrality Regulation*, Journal on Telecommunications & High Technology Law, October 2007.

[18] J. G. Sidak and D.J. Teece: *Innovation Spillovers and the 'Dirty Road' Fallacy: The Intellectual Bankruptcy of Banning Optional Transactions for Enhanced Delivery over the Internet*, Journal of Competition Law and Economics, No. 6, pp. 521 - 594, September 2010.

[19] F. Schuett: *network neutrality: A Survey of the Economic Literature*. Review of Network Economics, No. 9, September 2010.

[20] G. Schwartz, N. Shetty, D.-M. Chiu: *Impact of QoS on Internet User Welfare*. In Internet Network Economics. December 2008.

[21] B. Ruedt: *UMTS: Swisscom blockt Dienst fuer Internet-Telefonie*. 20 Minuten Online, July 2005. Last retrieved 6.11.13 from http://www.20min.ch/tools/suchen/story/27383201

[22] S. J. Vaughan-Nichols: *Verizon: No free tethering for unlimited data plan customers*. ZDNET, last retrieved 14.11.2013 from http://www.zdnet.com/verizon-no-free-tethering-for-unlimited-data-plan-customers-7000001987/

[23] T. Wu: *network neutrality, Broadband Discrimination*, Journal on Telecommunications & High Technology. Law, No. 2, pp. 141 - 178, June 2003.

[24] C. Yoo: *Beyond network neutrality*, Harvard Journal of Law & Technology, No. 19, pp. 1 - 77, June 2005.

Chapter 5

Bitcoins - Hype or Real Alternative?

Daniel Reber, Simon Feuerstein

This chapter gives an overview of Bitcoins, a recently established popular crypto currency. It first explains the technical aspects of Bitcoin. In particular it introduces the blockchain as a decentralized ledger and Bitcoin mining as a transaction verification and recording mechanism. Then, different parts of the Bitcoin ecosystem are presented, followed by the financial aspects of mining. Next, the evolution of currencies and the economic implications of the Bitcoin design are analyzed. Finally, the political side is examined through the viewpoints of various stakeholders and, after a short case study, possible future developments are considered.

Contents

5.1 Introduction

Bitcoin is a decentralized virtual crypto currency which lately received increasing attention in the mainstream media. The original concept was proposed under the pseudonym Nakamoto in 2008 [9] describing the technical foundations of the concept followed by an implementation in early 2009. This paper provides an introduction into the topic of Bitcoins and its technical bases. The system architecture is explained with a focus on the blockchain and its characteristics. Afterward a short overview over the current Bitcoin ecosystem with a special focus on the topic of mining Bitcoins is provided. To fully understand Bitcoin an economic analysis of the system is undertaken and a stakeholder analysis including regulatory provisions of different countries are discussed. To illustrate possible illegal usage of the Bitcoin system, the Silk Road marketplace case is presented. To conclude the future of Bitcoins is discussed.

5.2 What Bitcoin Is And How It Works

Bitcoin is a decentralized virtual crypto currency. The main technical concepts used for the implementation are hashing for a proof of work, asymmetric cryptography for authentication and verification and merkle trees for efficient storage. The starting point for most users is to create a wallet by installing a Bitcoin client. A Bitcoin wallet enables the user to send, receive and store bitcoins. Each wallet consists of one or more Bitcoin addresses whereas each address has its own public-private key pair. Transactions are authorized through the private key of the sending address and their validity is verified through the network utilizing the public key. Since Bitcoin is organized as a peer-to-peer system, a central authority to legitimate money transfer or manage the balance of each wallet is missing. Instead all transactions are made publicly available through the the so-called ”blockchain”, Bitcoins distributed ledger. Through analyzing all transactions, the balance for each existing Bitcoin address can be calculated and the current location of all bitcoins verified.
The blockchain consist of single blocks each containing transactions from the last few minutes forwarded to the particular block creator and information about the previous block. At the moment the blockchain consists of approximately 270’000 blocks [34], has a size of 11.9 gigabytes [30] and handles between 40 and 90 thousand transactions each day [32]. Each peer downloads the blockchain and can therefore calculate the balance for each wallet there is. This is possible because Bitcoins are only issued through generating the blockchain and get assigned to an address via a transaction. The final amount of Bitcoins is limited to 21 million.
Since transactions are propagated through the internet in a peer-to-peer network, a single transaction may take a moment to be propagated to all nodes. In the meantime, the transaction issuing wallet could authorize more transactions and spend the Bitcoins on other goods or services (the so-called ”double-spending problem”). To decide which transaction is issued first, the blockchain employs a cryptographic timestamping principle. The next chapter will describe the blockchain in more detail and discuss some technical details.

5.2.1 Blockchain

As previously stated, the blockchain stores all transactions and ensures the order of those to prevent double spending. To establish a chronology of blocks, each block contains information about the previous block (block A) and again some new open transactions thus forming the next block (block B). The system is therefore able to guarantee, that transactions contained in block A occurred before the transactions in block B. To prevent

attackers from propagating a different successor block B1 to block A throughout the network, the generation of a block has to be computationally complex. With each block added after block A, the task of propagating an alternative successor to block A becomes increasingly difficult. This fact will become of special interest when discussing the security properties of the Bitcoin System in subsection 5.2.3. The computational problem has to be complex to solve, but easy to verify, since all nodes in the peer-to-peer network need to be able to perform the verification task.
A mechanism that satisfies these requirements is finding a specific hash for specific data input while manipulating a nonce. The Bitcoin system uses SHA 256 to hash all transactions in a block and this information is included in the block header. The block header consists of the version, an SHA 256 hash of the previous block, hash of all transactions, a timestamp, the current difficulty and the flexible nonce. The task is to find a hash with a certain number of leading zeros (defined by the complexity) while only manipulating the nonce. Since the nonce is only a 32 bit value two problems arise. First checking only around 4 billion possibilities is not that time consuming and secondly changing only 32 bit of the hash input will most likely not generate a hash in the desired format. To solve the problem of the small changeable input, the hash of all transactions is changed as soon as all possible nonces have been checked. This changes 32 bytes of the block header which makes it again possible to generate a hash in the desired format. The changed SHA 256 hash of all transactions results from changing the generation (reward) transaction. [21]
While the blockchain prevents double spending by ensuring the succession of transactions, it also is responsible for generating bitcoins. With every new block generated, a transaction (previously called generation transaction) of 25 bitcoins is granted to the node which generated the block. The amount of bitcoins per generated block used to be 50 but was halved in late 2012 [24]. The system is designed to halve the reward every 210'000 blocks [22] and a block should be created every 10 minutes. At this rate, the reward will be halved again in about 3 years. To ensure a steady supply of newly generated bitcoins the system is designed to generate a block every 10 minutes. Every 2016 blocks the system checks how much time was needed to generate the last 2016 blocks [23]. To cope with increasing or decreasing computation power in the network, the System is able to adjust the complexity of generating a block. If the blocks have been generated to fast, which means in less than 14 days (2016 blocks x 10 minutes), the complexity will be increased. If the blocks have been generated too slow, the complexity will be decreased. In the last few months, the complexity increased drastically roughly every two weeks. Figure 5.1 [19] shows the complexity since February 2013 as a red line. As the diagram shows, the complexity increased sharp especially since September 2013, indicating a rapid growth in network computing power. The business of calculating the blockchain, also called "mining", will be discussed in section 5.4.

5.2.2 History

Based on the paper of Nakamoto Bitcoin was implemented and the first block in the blockchain, the so called "genesis block" was generated. The first exchange started in July 2009 at an initial exchange rate of 1'309.03 bitcoins for 1 USD. On the 30th December 2009, almost exactly one year after the generation of the genesis block, the difficulty for generating a block increased for the first time. In July 2010 the still existing Bitcoin exchange MtGox started its operation. In September 2010 the reward for a block was split for the first time, thus introducing the concept of mining pools. In November 2010 the market capitalization for the all generated Bitcoins reached 1 Million US-Dollars. During the year 2011 several Bitcoin exchanges started operation and the exchange rate climbed from 1 USD per Bitcoin in February to 10 USD in June and ended up being around 4 USD in December 2011.

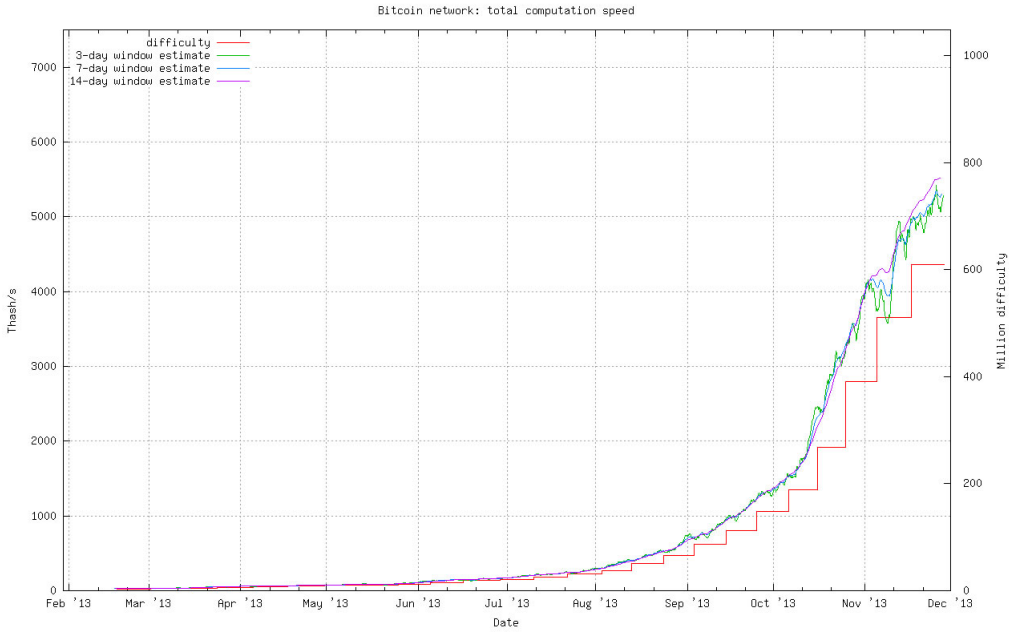

Figure 5.1: Development of block calculation complexity

The year 2012 brought the halving day on November 25th when the reward for generating a block was halved from 50 bitcoins to 25. An Increasing number of transactions were caused by one service called SatoshiDICE, a gambling service with instant payout. In 2013 the market capitalization of all bitcoins reaches 1 billion USD on March 28th and almost reaches 10 billion USD on November 24th (see 5.2). During 2013 Bitcoin also receives more attention from Mainstream media as well as governments which results in first regulations, for example in the US [24].

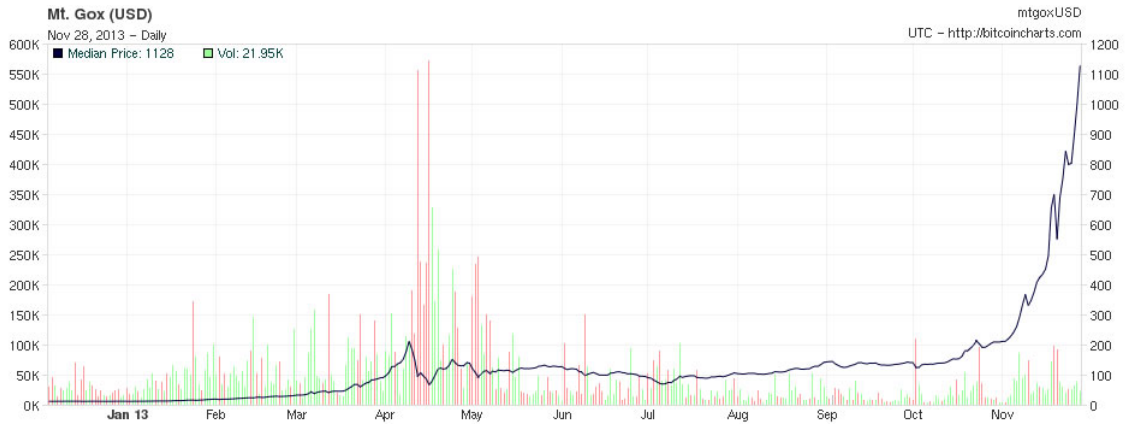

Figure 5.2: Exchange Rate Development

5.2.3 Anonymity and Security

Since the blockchain makes all transactions public, privacy issues might arise. The Bitcoin Project site [17] proposes some measures for protecting one's identity. To receive goods or services, you usually need to reveal your identity so if a user utilizes a single address to send and receive payments, his identity is known by the counterpart. To avoid traceability, the bitcoin creators suggest three simple measures. First, one should always use a new address to receive payments which prohibits linking multiple transactions to one address. Second, address changes should be used when sending payments so the old address has a balance of zero and it is unclear which address actually received the payment. And as

the third measure, the own address should not be published since this makes linking the address to a real person easy.
Even if a user takes such measures, anonymity is not guaranteed. Meiklejohn et.al [6] demonstrate how transactions of stolen bitcoins can be traced back to individuals. But there are already proposed improvements like zerocoin [56] that provide additional security (see also subsection 5.8.2).
Besides Anonymity, also the security of the network is an issue. The power to create an alternative fork in the blockchain enables changes according to the attackers needs (so-called "history-revision attack"). As the original paper states [9] at least 50 percent of the networks computation power need to be honest. Nakamoto argues, that the majority will always outpace the attackers, since they have more computation power than the attackers. Because the network will always adopt the longest blockchain, an attacker would need more than 50 percent of the computation power. On the other hand Eyal and Sirrer [4] argue that already 1/3 of computing power or even no lower threshold at all allows to attack the blockchain. They argue that selfish miners will form a group and keep generated blocks private, as soon as the attackers are able to mine several blocks in a row, they can publish them successively when the public chain increases its length. This renders the efforts of all other miners useless and increases the return for the selfish miners. Evidence for the possibility of such attacks comes from the fact, that the mining Pool BTC Guild was able to mine 6 consecutive blocks in 2013 [46]. On the other hand, renowned scientists like Harvard Professor Ed Felten doubt that the attack described in the paper is possible [16]. The main point he rises is that keeping the blockchain private is impossible since the so called fair-weather miner would always mine on the longer blockchain, thus rendering the attack impractical.

5.3 The Ecosystem

This section dives into Bitcoin's ecosystem, which is already remarkably diverse for a system this young and keeps on growing rapidly. The section starts with some general statistics of Bitcoin's current size, followed by a look at its usage and real life spending possibilities. Finally, several services are introduced to provide some idea of Bitcoin's diversity.

5.3.1 Statistics

At the times of writing (November, 26, 2013) Bitcoin's exchange rate has climbed to an all time high of 1210.00 USD rising from around 150.00 dollars two months ago. 12 million bitcoins have been minted so far, of which an estimated 4 million are circulating and 8 million are being hoarded with a hard to estimate reminder of lost bitcoins. The dominant usage is gambling which makes up around 64% of the daily 100'000 transactions [14, 6]. The complete blockchain since the genesis block is a considerable 11.9 GB [31].

5.3.2 Goods Market

Bitcoin is used both by legitimate online market places and illegal ones, like the now defunct Silk Road. Under the early adopters you can find quite a few web design companies, hosting and domain services. Most often Bitcoin is used as an alternative payment method alongside traditional options like PayPal or credit card transactions. Examples for Bitcoin usage in Switzerland are Tazzine, an online shop for coffee mugs that accepts bitcoins [47]; bitcoinstore, a electronics vendor [18] or bitmine, a seller of mining equipment [27].

5.3.3 Services

A flourishing service industry has established itself around Bitcoin. The most important service are currency exchanges like MtGox [44] and BitStamp [28]. For information about these markets statistics sites like bitcoincharts [14] that provide up to date market data for all major exchanges and useful information like the current hashrate distribution between largest mining pools exist. For the inspection of the blockchain interested user can use blockchain.info or the block explorer [29, 34] which enables the analysis of specific transactions and the tracking of bitcoins through the blockchain. New sites like coindesk [37] keep readers up to date with the most recent developments in the Bitcoin world. Finally, various message boards and blogs of miners, scientists, developers and journalists assure the continuous exchange and dissemination of ideas. The most important forum is bitcointalk [20].

5.4 Bitcoin Mining

As previously stated mining is the task of generating the blockchain and securing transactions. The reward for mining is 25 bitcoins per found block. At current exchange rates, this translates to 25'000 USD per found Block. Mining is competitive, the miner that finds the block first gets the reward and all other miners don't receive anything. Depending on the mining equipment and the complexity, finding a block can take years. To make the rewards more steady, miners form groups, so called mining pools. At the moment about 85 percent of the whole network calculation power is organized in pools [33]. The two biggest pools, BTC Guild and GHash.IO, account for 57 percent of the overall computing power. Those pools split up the work of generating a block and then split the block reward according to the contribution of each pool member [26].

5.4.1 Evolution of Mining Equipment

The performance of mining equipment is measured in million hashes per second (Mhash/s). Since the introduction of Bitcoin, several mining technologies have been used for generating the blockchain. In the beginning standard CPUs were used to create SHA-256 hashes. Up to date devices, like an Intel Core i7 are capable of mining at a speed of 19 Mhash/s while consuming 150 watts of power. The first increase of mining speed was reached through utilizing GPUs for generating the hashes. Up to date GPUs like an AMD Radeon 7970 are capable of mining at a rate of around 800 Mhash/s while consuming 200 watts of power [25]. The second improvement was reached with the introduction of FPGAs for mining. While the computation speed remained similar to GPU's the effectiveness increased resulting in lower power consumption. The latest evolutionary step in mining hardware are ASIC chips. Those chips are optimized for the generation of SHA-2 hashes and are therefore a lot more efficient than any technology used for mining before. The performance of such devices reaches from around 5000 Mhash/s for the butterfly labs 5GH/s Bitcoin Miner [35] to over 1'000'000 Mhash/s for the Bitmine.ch 1 TH/s [36]. Since the technology is pretty new and, due to the high exchange rates, profits are pretty substantial, the demand for the devices is high, thus making the delivery date for the individual very uncertain.

5.5 Calculated Scenarios

When calculating scenarios, there are several factors that influence the profitability of a mining operation. The first and most unpredictable is the exchange rate of bitcoins. The

second factor is the delivery date of the hardware and connected to this, the complexity at the time of the delivery. Since the difficulty increase is not predictable and dependent on the network computation speed, estimating a complexity at a given point in time in the future is almost impossible. For the scenarios displayed in table 5.1 a complexity growing rate of 1.5 in the optimal scenario and 1.7 for the conservative scenario were chosen while leaving the exchange rate at the current high of 1'000 USD per Bitcoin. For calculating the scenarios a Bitmine.ch Coincraft 1 TH/s for 5'600 USD and consuming around 900 watts of power was chosen [36]. At the time, the delivery date was estimated in the 3rd week of January, leaving a 60 day windows until delivery.

Table 5.1: Scenarios

	Optimal	Conservative	Falling Exchange Rate
Growing Rate	1.5	1.7	1.5
Complexity	6'942'388'651	14'711'430'242	6'942'388'651
Exchange Rate	1000 USD / BTC	1000 USD / BTC	250 USD / BTC
Break-even	86 Days	208 Days	>10'000 Days
Return 1 Year	14'092 USD	3'068 USD	-3'957 USD
Return 2 Years	25'643 USD	7'894 USD	-3'874.92 USD

The different estimated complexity growth rates in the first two scenarios show that only small changes in the growth rate can have big impact on profitability. While the optimal scenario reaches break-even after 86 day, it takes around 200 in the second scenario. Assuming a falling exchange rate in the third scenario with optimal complexity growth, the break-even would be reached in more than 10'000 days. Given the high uncertainty in estimating the variables that lead to the scenarios above, mining bitcoins is a risky endeavor.

5.6 Economics of Bitcoin

In this section Bitcoin is examined from an economic perspective. It begins by looking at what kind of currency Bitcoin is and where it fits in historically followed by an overview of the most important characteristics of the Bitcoin system in its current incarnation.

5.6.1 Functions of Money

Modern economists see money performing three distinct functions. First, it is a medium of exchange. This functions as a solution to the "coincidence of wants" problem which arises in moneyless barter systems when a market participant has to find a trading partner who not only has what the first participant wants, but also coincidentally wants what he has to offer in exchange. Second, money facilitates the comparison of the value of different goods by functioning as a unit of account. And third, it is a store of value that enables long-term planning and behavior. Bitcoin fulfills the functions of unit of account and medium of exchange and as a digital currency, some might argue, even better as paper money since it enables non-local transfers without additional transaction fees. Whether Bitcoin is also a good store of value is currently contested. Bitcoin advocates see them as superior because governments cannot tamper with their value through printing money or asset seizures like the recently proposed 10% cut on savings accounts during the Cyprus crisis. Opponents point to the rampant exchange rate volatility [2, 10].

5.6.2 Historical Context

Historically there has been a progression from more physical towards more abstract forms of money. Early forms of commodity money like cattle or seeds had intrinsic value. Commodity money based on metals derived their value primarily from their scarcity, especially if one thinks of jewelry as a form of wealth storage and not intrinsically valuable. The next logical step was a shift to representative, commodity-backed money to avoid the dangers and logistic challenges of large metal transfers. This form of money was in principle redeemable for the underlying commodity (Gold certificates are a good example). Modern fiat money drops this requirement. In its place steps the trust in the issuing government who gives paper money value through legal decree [3].

5.6.3 Modern Currencies

With the advent of modern communications systems fiat money received an upgrade to incorporate the advantages of fast transfer across large distances, nowadays in the form of the digital commercial banking system and hereby turned even more abstract from paper to bits and bytes. The internet finally brought digital money to the customers but electronic money services like PayPal remain fully linked to real fiat currencies by being quoted and valued one to one to the denomination of the PayPal account. Like in the commercial banking system, an electronic PayPal US dollar equals a physical US dollar. Virtual currencies are the newest innovation in the evolution of money. The European Central Bank (ECB) distinguishes between three types of virtual currencies. Closed schemes where participants both receive and spend the currency inside its ecosystem, like Blizzard Entertainment's World of Warcraft Gold. Unidirectional schemes where you acquire the virtual currency in exchange for real money but can spend the virtual money only inside the limits of its scheme, like Microsoft points or frequent-flyer miles. And finally bidirectional virtual currencies like Second Life's Linden dollars or Bitcoin [3].

5.6.4 Intellectual Heritage

Bitcoin's design has its roots in libertarian ideas in the tradition of Austrian economist like von Mises and Hayek. The Austrian school believes that state respectively central banks are tempted to abuse their control over their fiat currency through expansionary monetary policy for dept financing or artificial growth. The fractional reserve banking system intensifies the money creation and leads to exacerbated business cycles and inflation. Inflation has several negative effects that distort an economy: The opportunity costs of holding money rise, more trips to the bank have to be made (so called shoe-leather costs), businesses have increased menu costs (the frequent readjustment of advertised prices), more taxes have to be paid since both income and capital gains taxes are based on nominal valuation, wealth is redistributed from creditors to debtors since the original credit is worth less, individual decision making becomes more difficult (money illusion) and existing capital and savings get devaluated [2]. If not brought under control, these distortions can have severe consequences like mass unemployment, political unrest and famine and even the total collapse of a national economy.
The remedy Austrian economist prescribe is the abolition of central banks and a return to the gold standard, which is harder to manipulate because of the inherent scarcity of gold. Bitcoin's two main economic features are its fixed money growth via the mining mechanism and, owing to its peer to peer design, no central authority that could influence the currency. These features mirror the Austrian approach. Despite the ideological proximity modern libertarians criticize Bitcoin's lack of intrinsic value, even thou the intrinsic value of gold is not quite undisputed too (It fails the "What would take with you to a lonely island?"- test).

George Selgin argues accordingly that money should not be split along commodity/ fiat lines, but proposes a classification along the dimensions of scarcity (absolute/contingent) and nonmonetary use (yes/no). Bitcoins with its absolute scarcity and no nonmonetary use would be considered "synthetic commodity money" in this classification [11].

5.6.5 Implications of the Design

The economic consequences of Bitcoin's design are an intensively debated subject in the Bitcoin community which, due to the recent spike in the exchange rate, attracts more and more attention from the media as well and has motivated initial scholarly examinations from researchers. Generally it is important to differentiate between short-run and long-run effects. Short-run problems will arise during the adaption of Bitcoin and concern issues like its adaption, regulatory reaction and impact on existing currencies. Long-run problems arise when a Bitcoin economy is fully established and the limits of its design features come apparent. One such long-run problem is deflation.

5.6.6 Threat of Deflation

The anti-inflationary mining mechanism is a double-edge sword, because it leads to strong deflationary tendencies. Bitcoin in its end state will have a fixed monetary base facing a growing amount of goods. This implies falling prices per good, which leads consumers to delay consumption since they will have a higher future purchasing power. That shift in the demand curve leads to even lower prices in the short run and triggers a deflationary spiral. At the current stage, with a still expanding monetary base, this is not yet highly relevant. It seems sensible to attribute the current form of hoarding to short to mid-run speculation and not to bets on lower prices because the goods market is still in its infant stage [1, 3].

5.6.7 Mining

Section 5.4 examined the financial aspects of Bitcoin mining and the volatile factors that influence the potential payoff. This section focuses on the long-run incentive structure miners have. The block-chain is the backbone of the Bitcoin system. To guarantee the sustained functioning of Bitcoin it is therefore of utmost importance that the incentives of miners are congruent with the intended protocol. The absence of a competitive and honest mining community lies at the heart of many Bitcoin design pathologies. Kroll, Davey and Felten have written the first paper to analyze the economics of Bitcoin mining [5].
A potential pathology they identify is the possibility of a death spiral. One of the things the previous examination of mining in section 5.4 has shown, is that the exchange rate is the essential volatile factor in mining investment calculations. A lower exchange rate could render an investment unprofitable. The death spiral occurs when a loss of confidence in Bitcoin leads to lower demand for bitcoins and an increased supply on exchanges by users wanting to flee the currency. This could rapidly lower the exchange rate thus disincentivize miners which in turn makes the system more vulnerable to a history-revision attack which lowers the confidence in the system further thereby starting the next round of the spiral. The difficulty of a history-revision attack is a contested matter. The Bitcoin design assumes that 51 percent of mining power is the necessary threshold for a sustained alternate fork to prevail, 30% and 25% and no lower bound at all are proposed as well [4] As mentioned in subsection 5.2.3, the mining pool BTC guild for instance has already mined six blocks in a row [46]. But the capability to execute such an attack does not necessarily imply that it is worth doing so. If one assumes that a single large revision or several

medium revisions destroys the confidence necessary for the continuation of Bitcoin, then attackers need to gain enough to cover their technical investment plus the devaluation of their remaining bitcoin holdings, which seems highly unlikely. But Eyal and Sirer [4] show that the incentives for a form of sustainable forking are given. The authors describe selfish mining that aims not at revising the block-chain history for double-spending purposes but to gain a disproportionate mining reward in relation to the computation power contributed. They show that a mining pool of any(!) size can gain through selfish mining and will attract a majority of miners, leading to monopolization. Next to gaining control of the network for double-spending or coin rewards, a third motivation exists: the destruction of Bitcoin as the goal itself. Law enforcement, activists or short sellers could see gains in the destruction of Bitcoin [5] Last but not least block creation control gives the power to select or ignore the transactions to be executed.
Overall four factors pressure the system towards mining monopolization: Frist, the technological race for more hashing power raises the bar for the entry of new miners. Second, the possibility of selfish mining incentivizes miners to join such a pool for higher rewards. Third, the decreasing bitcoin yield discourages the entry of new miners in the long run. And fourth, the declining reward should be offset through an increase in transaction fees to compensate miners for their effort. But substantial transaction fees will only establish themselves if monopolization has already taken place. In the case of a diverse and competitive mining community downward pressure should keep transaction fees low, since from the perspective of a miner every non-zero fee not included in its block could be claimed by another miner. A fifth potential factor would emerge, if miners will be classified as Money Transfer Businesses (MBTs) with all the corresponding legal requirements which would basically prohibit regular persons from mining.

5.6.8 Further Consequences

Another question is if Bitcoin is a Ponzi scheme. The question is not about Ponzi schemes conducted inside Bitcoin but the overall design. Because Bitcoin shares, in the absence of a mature goods market where you could liquidate your holdings through consumption, a Ponzi scheme characteristic. Namely, to leave Bitcoin you have to find a new participant willing to purchase your bitcoins [3]. And finally the question remains, whether the explosion in the exchange rate reflects some underlying value or is mainly a speculative bubble. As mentioned in subsection 5.3.1, the largest part of transactions inside the Bitcoin economy stems from gambling. As long as the goods market does not accelerate its growth significantly there is no reason to assume it is not a bubble. New entrants meet an illiquid market which pushes up the exchange rate. If there was a significant amount of vendors who regularly converted their Bitcoin earnings into cash, there would be at least some downward pressure on the exchange rate.

5.7 Politics

This section concerns the actors who have a stake in the future development of Bitcoin. In addition to customers and businesses, there are powerful actors like central banks, international organizations and nation states. But first this section begins with an general overview of Bitcoin's inherent risks.

5.7.1 Kinds of Risk

Bitcoin poses several risks: first, credit risk, the risk that participating parties cannot meet their financial obligations; second, liquidity risk, the risk that a user cannot access

or exchange bitcoins; third, operational risk, the risk that arises through the reliance on the functioning of the Bitcoin network; fourth legal risks, risks stemming from legal uncertainty that amplify other risks and finally the risk of outright fraud [3].
Tyler Moore and Nicolas Christin published a paper that illustrates the riskiness of Bitcoin in the case of exchanges [8]. They conducted a survival analysis based on a dataset of 40 Bitcoin exchanges during a three year period. From these 40 exchanges 18 had closed, only 6 of which seem to have reimbursed their customers. Average daily transaction volume, the experience of a security breach and financial regulation compliance of the host country (based on a World Bank index) were used as independent variables based on the following hypotheses: (1)Larger transaction volumes should correlate with higher profitability which assures the continuation of operations. (2) A security breach diminishes both profits and customer confidence, increasing the probability of a shut down. And (3) stricter compliance with anti money laundering and combating the financing of terrorism regulations should increase the risk of closure. The best-fit Cox model showed only a significant coefficient for the transaction volume (-0.173, $p = 0.016$) but the experience of a security breach had a strong effect size (0.857), its non-significance stems from the limited sample. The regulatory index does not seem to capture the willingness to apply financial regulation towards Bitcoin. This is to be expected at this point in time where most countries lack experience with Bitcoin related cases. Overall both hypothesis (1) and (2) found support. Figure 5.3 shows the survival probability functions for several exchanges. Note how even for the well established exchange MtGox (green) the 800 day survival probablity drops to around 60 percent.

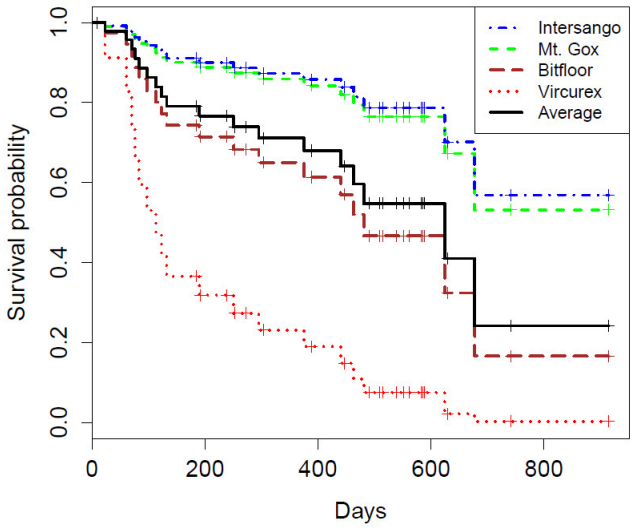

Figure 5.3: Survival Probability Functions for Bitcoin Exchanges

5.7.2 Customers

Individuals value the anonymity, independence and the ease of payment the Bitcoin infrastructure provides. Be it as commercial customers or as private citizens as in the case of remittances, (international money transfers from immigrants back to their home country), that are getting more difficult via conventional channels due to increased legal scrutiny in the wake of anti-terror legislature [48]. The anonymity can be valued both by legitimate and illegitimate users to stay out of reach of an oppressive government or to avoid regular law enforcement. But from the risks mentioned above several apply to individual customers. In particular liquidity risks, legal uncertainty and fraud are of im-

mediate concern. These risks could incentivize the acceptance of potential regulation or governance.

5.7.3 Businesses

The business side profits currently from minimal transaction cost and the irreversibility of transactions, a constant concern with traditional payments and services like PayPal where defaults on obligations and repossessions interfere with the normal business process. There are many opportunities for both new and established businesses. Most of financial engineering products and services used with regular markets for instance can be ported to the Bitcoin world. The main concern for businesses is the legal uncertainty they have to cope with when dealing in bitcoins. Additionally they face the same kinds of risks as the customer side. A large multinational cooperation as an advocate would help Bitcoin immensely both in terms of reputation and legislative bargain power.

5.7.4 Central Banks

A central banks responsibility is the monetary policy of a nation. The goal of central banks is low inflation, low unemployment with accompanying high growth. To influence the economy accordingly they nowadays target short-term interest rates by controlling the money supply mainly through open market transactions and the influencing of market expectations. As mentioned in subsection 5.6.4, Bitcoin wants to make central banks obsolete by replacing their role as money issuing institution through a decentralized alternative. Substitution of a central bank's own currency with bitcoins diminishes the central bank's control of the money supply and therefore its influence over interest rates. In the long run this poses potentially a direct threat to the existence of central banks and worries their advocates who see them to be fulfilling an important function in a modern economy [3].

5.7.5 International Monetary Fund

The international organizations that is interested in Bitcoin the most is the International Monetary Fund (IMF). The IMF, together with the WTO and the World Bank, are the cornerstones of the contemporary international economic regime. The IMF's role is to guarantee exchange rate stability and the prevention of speculative attacks on national currencies. This makes the IMF *the* international body the most affected by Bitcoin. To achieve this goals member nations supply the fund with a quota of their national currency in exchange for regulated drawing rights in case of an emergency. Bitcoin has no nation behind it, which leaves the IMF short of Bitcoin reserves. If Bitcoin became an established currency there would be no possibility to counter a speculative attack. An IMF report suggests two ways to ameliorate this situation. Either the fund receives bitcoins indirectly by requiring its member nations to deposit a part of their quota in bitcoins or directly by granting certain digital currencies membership status, enabling users to deposit bitcoins directly with the fund. The later proposition would probably meet both ideological and motivational difficulties [10].

5.7.6 Nation States

The behavior of nation states will be critical for the trajectory of Bitcoin in the foreseeable future. Governments have a wide array of interests that conflict with Bitcoin. First, states want to be able to tax economic behavior inside their territory. Different sorts of taxes could apply to Bitcoin depending on the legal classification. Is it considered a good,

value added tax (VAT) would apply or is it rather considered a form of currency or financial instrument, capital gains taxes would come into play. Second, states want to support the smooth functioning of commercial interactions by providing legal security. The irreversibility of Bitcoin transactions as an example poses a difficulty to contract law. Third, governments want to enforce their criminal law code be it regarding the trade of illicit goods and services, money laundering or the financing of terrorism.
To achieve this existing legislation has to be modified or new law dealing with digital currencies has to be written. Several countries are in the process of educating themselves about Bitcoin, Switzerland included [45]. Some have already taken first steps to assure a minimum of legal guidance until further legislation can be drafted. Germany classifies bitcoins as a form of private money and treats it similar to other financial instruments like stocks. Long-term holdings do not have to pay capital gains tax at the moment. Thailand was reported to be the first country to ban Bitcoin, what turned out to be only a review of a license for a single Bitcoin exchange. The United Kingdom classifies bitcoins as vouchers so that in principle VAT should be paid. In the United States different states treat Bitcoin differently, due to the common law practice, where until supervening federal legislation is passed state judges rulings determine the momentary legal standing. Some bank accounts associated with bitcoin transactions have been frozen. Exchanges are considered MBTs (see subsection 5.6.7) and have to comply with respective laws. The legal status of miners is currently unclear, but there are arguments to classify as MBTs as well, which would have far reaching consequences for the proper functioning of the Bitcoin system [38, 39].

5.7.7 Law Enforcement

Bitcoin's anonymity poses an obstacle to tax departments and law enforcement agencies. But as a public ledger the block chain can be used to track the history of every bitcoin ever mined. Known fraudulent bitcoins can therefore in principle be followed until they try to exit the system through an exchange, where you can subpoena information about the corresponding account holder. The exit and entry points between the Bitcoin world and different systems (exchanges, markets, message boards etc.) are generally the points of potential exposure. Graph mapping and statistical analysis can be used to associate different public keys with a user to help with this identification. Mixing services or secondary layers of anonymization, like the Tor network, make this endeavor more difficult but not impossible [6, 54]. Large scale identification of participants on the other hand seems out of reach for even the most sophisticated entity.

5.7.8 Case Study: Silk Road

The Silk Road case serves as a good illustration of the criminal usage of Bitcoin. Founded in February 2011 by the user "Dread Pirate Roberts" (DPR), who later turned out to be 29 year old Ross Ulbrich [50]. Silk Road was an online black market for legal and illegal goods. Even though the terms of usage suggested to trade only goods that "do no harm" like drugs and prohibited the trade in child pornography, weapons or weapons of mass destruction, one could find stolen credit cards, banking Trojans and similar illegal goods and services. During its existence Silk Road amassed 1.2 billion dollars in revenue and 79.8 million dollars in commissions and had 146'946 buyers and 3'877 vendors. Silk Road users connected through the Tor network to the site and paid their purchase with bitcoins [53, 42].

5.7.8.1 Investigation

Silk Road attracted a multi-agency task force that "included investigators from the FBI, DEA, DHS, the IRS, U.S. Postal Inspection, U.S. Secret Service, and the Bureau of Alcohol, Tobacco, Firearms and Explosives" [51] The fist breakthroughs were multiple arrests of vendors mainly due to negligence (e.g. leaking packages, previous criminal history) that allowed the agents to gain access under their identities to get nearer to DPR. The next step was a sting operation that led to a Silk Road administrator. DPR found out about the investigation against the administrator and hired an undercover agent to kill the compromised admin. The agents staged a mock killing and sent pictures as proof. Meanwhile the investigators could link a forum account from during the genesis of Silk Road to a Gmail account. This provided enough information to gain access to Silk Road's hosting server. Either poor encryption or the seizure of a stored key seems to have helped the investigators. An interception of fake IDs to Ulbricht's home that he had ordered was the final step [51, 43].

5.7.8.2 Ramifications

Illustrating the illegal usage of Bitcoin aside, the take down of Silkroad serves as a natural experiment for the resilience of the Bitcoin economy. October second trading data shows a sharp decline in the Bitcoin/USD exchange rate from approximately 126 dollars to around 85 dollars probably caused by Silk Road vendors who panicked and tried to liquidate their remaining bitcoin holdings (see Figure 5.4). The seized bitcoins from the accounts were obviously inaccessible. Until October fifth the exchange rate had already stabilized on its previous level and stood at the date of writing at over 900 dollars (see Figure 5.5). Proponents of Bitcoin see this to indicate the legitimacy of Bitcoin. In their argument illegal activities are not the driving factor of Bitcoin usage and the confidence in Bitcoin remains strong [40]. Opponents could see it as a further confirmation for a bubble, since at least bitcoins were used for the trade in goods. The remaining real trade remains miniscule compared to gambling and hoarding.

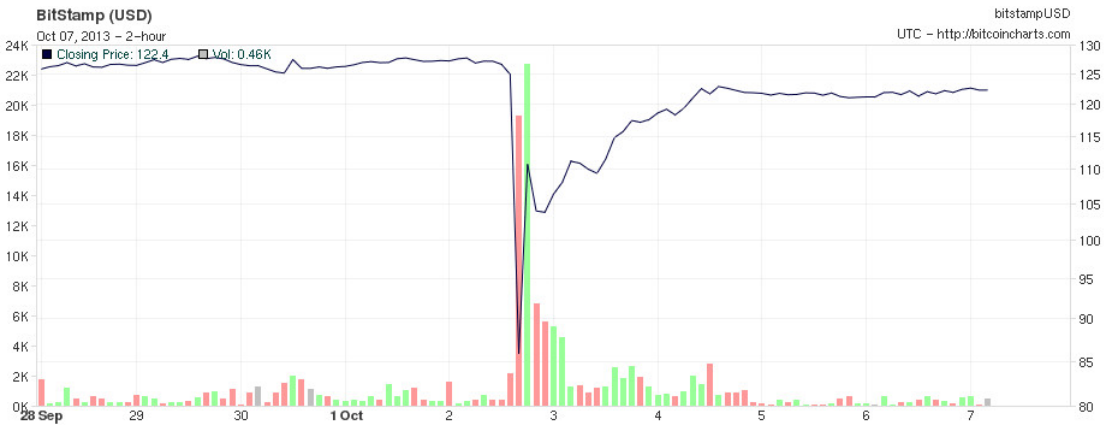

Figure 5.4: Impact of the Silk Road Arrest

5.7.8.3 The Legacy of Silk Road

The law enforcement side is rid of a public embarrassment and could show that it is capable of bringing an investigation against a Bitcoin based criminal endeavor to completion but Ulbricht's security negligence was a major contributing factor. How well law enforcement fares against a proper implementation of a Bitcoin black market remains to be seen, especially since the criminal side will learn its lessons from this failure and will evolve. Silk Road 2.0 implementations are already up [41].

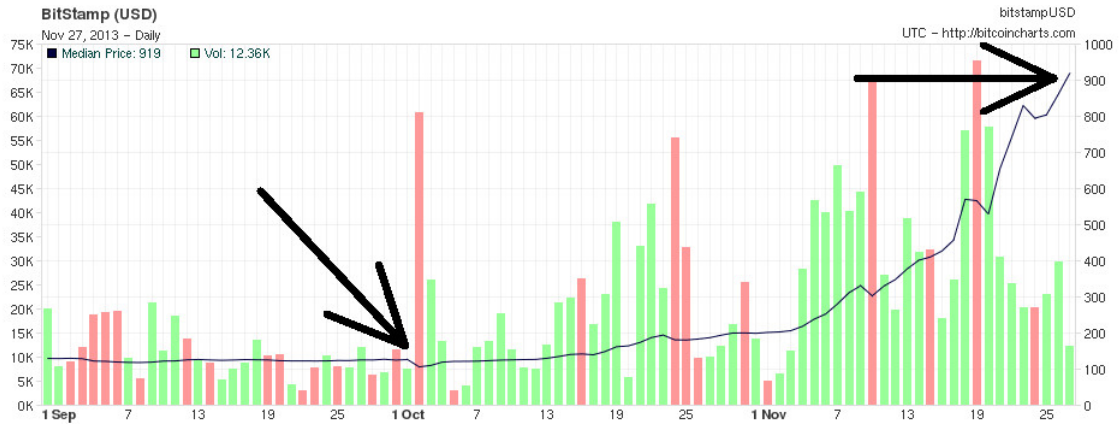

Figure 5.5: Long-term Effects of SR Arrest

5.8 The Future of Bitcoin

This final section looks at Bitcoin's possibilities to develop. Like with any emerging field many ideas are proposed and ready to be tested. Some ideas are already in the process of implementation and others only exist as proposals from researchers and developers so far.

5.8.1 Technical Future of Bitcoin

On the engineering side Bitcoin can take three general directions. The Bitcoin design can be extended, changed or be replaced by completely new designs. Extensions build layers on top of the Bitcoin architecture and take advantage of the underlying public ledger to accomplish their desired task. Examples are virtual notaries or bitbonds [13]. Somewhere between an extension and a change is Zerocoin. Zerocoin improves Bitcoin by providing strong anonymity. Zerocoin is cryptographically more advanced than Bitcoin but uses Bitcoin as the underlying currency. To simplify the description of its inner workings it is possible to see Zerocoin as a form of mixing service. Users can cryptographically commit their bitcoins to the Zerocoin pool, gaining a form of virtual credit. If the user wants to retrieve coins, a public accumulator verifies the existence of the committed coins and grants *any* bitcoins in the specified amount from the pool. This breaks the association between coin and user the same way a mixing service does [56, 7].
Barber, Boyen, Shi and Uzun [1] published a paper outlining several improvements to the Bitcoin design. The deflationary nature could be resolved by changing the change in money supply to linear money growth which guarantees a small constant inflation or even to dynamic money growth that is adjusted to the transaction volume to reflect real economic activity. As a guard against history-revision attacks the authors propose checkpointing, where users keep a cryptographically tamper-proof record of the block chain history so that they can spot drastic changes and communicate their disapproval to the network whereby older record holder would be given more weight than younger ones. The hard-coding of certain blocks during new Bitcoin client releases is a minor form of checkpointing that takes place today. To simplify the backup process pseudo-random generated keys could be used so only the seed would have to be stored or an expiration date of public keys would allow the user to discard obsolete private keys. Barber et al. further suggest a subscription based filtering service against network overload. For block generation and verification miners want to receive all possible packets. Accounts of simple customers on the other hand do not need most of the peer to peer traffic. A cloud based filtering service could filter packets and forward them to subscribers. There could be negative consequences of the introduction of critical nodes.

5.8.2 Economic and Political Future of Bitcoin

The economic and political future is very uncertain but could potentially interfere with the plans of the Bitcoin engineers and participants in Bitcoin related projects. As section 5.6 has shown, are the economic incentives not as intended. It will be interesting to see if the predicted mining monopolization occurs and if the bursting of the bubble or growing security concerns will undermine confidence in the system. And if and which changes to the Bitcoin protocol will consequently be made to counter such negative trends.
Politically it seems clear that some form of oversight is on the horizon. Governments simply cannot afford to ignore Bitcoin and virtual currencies in general. Taxation will likely be the first form of intrusion. If nations are unable to control Bitcoin, be it in its contemporary form or with extensions and changes like Zerocoin, they will try to outlaw or disrupt Bitcoin. One could point to the failed attempts to combat BitTorrent and similar file sharing programs and argue that such measurements will be futile but first, a virtual currency relies much more on confidence. Second, with the mining mechanism Bitcoin provides a clear point of attack. And third, governments' motivation is several magnitudes higher when it comes to tax evasion and especially potential financing of terrorism, than in the case of property rights. A possibility is the evolution of the Bitcoin Foundation from low key quasi governance to something more like the oversight of an open source project or even more formalized to offer governments a counter party in regulatory negotiations and disputes [1].

5.8.3 Conclusion

On the technical side the concept of Bitcoin appears to be stable and capable of handling an increased amount of transaction. Questions about the anonymity in the network are raised, especially when more addresses can be linked to an identity. Concerning the blockchain, the the minimum amount of "honest" computing power required for it to be safe, is up for debate. The generation of the blockchain, seems to attract lots of players due to the high exchange rates which reward miners with 25'000 USD per block. The entry of new players with specialized equipment leads to a sharp increase in the computation power available in the network thus increasing the blockchain complexity. From an economical viewpoint the Bitcoin System remains questionable since the amount of bitcoins is fixed which may lead to a deflationary spiral. To reduce the risk of deflation, several strategies to put a "Bitcoin Reserve Bank" in place are discussed. The political actors try to prevent felonies like money laundering, murder for hire and terrorism as well as offences like drug dealing or tax evasion. Therefore a lot of nation states are considering regulatory measures against Bitcoin. In the end, a lot of the attention Bitcoin attracts from different parties stems from the increased exchange rate. Through the mainstream media attention and announcements from big players to accept bitcoins as alternative payment method, the exchange rate is pushed further up. This behavior is a clear indicator for a bubble that will burst sooner or later. In summary, Bitcoin serves as an interesting case study to observe what difficulties and obstacles a distributed virtual currency faces. Even if Bitcoin in the long-run would not prevail it certainly generated valuable lessons for future research and development.

Bibliography

[1] S. Barber, X. Boyen, E. Shi, and E. Uzun. Bitter to Better – How to Make Bitcoin a Better Currency; (Financial Cryptography and Data Security), March, 2012, p. 399-414. http://crypto.stanford.edu/~xb/fc12/bitcoin.pdf.

[2] O. Blanchard: Macroeconomics; Book, Prentice Hall, 2008.

[3] ECB - Virtual Currency Schemes; http://www.ecb.europa.eu/pub/pdf/other/virtualcurrencyschemes201210en.pdf, November, 2013.

[4] I.Eyal and E. G. Sirer, (2013). Majority is not Enough: Bitcoin Mining is Vulnerable; arXiv preprint. http://arxiv.org/abs/1311.0243.

[5] J. A. Kroll, I. C. Davey, and E. W. Felten:The Economics of Bitcoin Mining, or Bitcoin in the Presence of Adversaries; Proceedings of WEIS 2013, June, 2013, http://www.weis2013.econinfosec.org/papers/KrollDaveyFeltenWEIS2013.pdf.

[6] S. Meiklejohn, M. Pomarole, G. Jordan, K. Levchenko, D. McCoy, G. M. Voelker, and Savage, (2013, October). A fistful of bitcoins: characterizing payments among men with no names; (Proceedings of the 2013 conference on Internet measurement conference), October, 2013, pp. 127-140 http://www.umiacs.umd.edu/~tdumitra/courses/ENEE759D/Fall13/papers/Meiklejohn13.pdf.

[7] I. Miers, C. Garman, M. Green, and A. D. Rubin: Zerocoin: Anonymous Distributed E-Cash from Bitcoin; (Proceedings of the IEEE Symposium on Security and Privacy 2013), May, 2013, pp. 397-414. http://spar.isi.jhu.edu/mgreen/ZerocoinOakland.pdf.

[8] T. Moore and N. Christin: Beware the Middleman: Empirical Analysis of Bitcoin-Exchange Risk; (Proceedings of Financial Cryptography 2013), April, 2013, pp. 25-33, http://fc13.ifca.ai/proc/1-2.pdf.

[9] S. Nakamoto: Bitcoin: A Peer-to-Peer Electronic Cash System; paper, May, 2009, 9, http://bitcoin.org/bitcoin.pdf.

[10] N. Plassaras : Regulating Digital Currencies: Bringing Bitcoin within the Reach of the IMF; (Chicago Journal of International Law, 14) , April, 2013, pp. 387-91, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2248419.

[11] George Selgin - Synthetic Commodity Money; http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2000118, November, 2013.

[12] Analysis of Silk Road's Historical Impact on Bitcoin; http://thegenesisblock.com/analysis-silk-roads-historical-impact-bitcoin/, November, 2013.

[13] Bitbond; https://www.bitbond.net/, November, 2013.

[14] Bitcoin Charts; http://bitcoincharts.com/, November, 2013.

[15] Bitcoin Foundation; https://bitcoinfoundation.org/, November, 2013.

[16] Bitcoin isn't so broken after all; https://freedom-to-tinker.com/blog/felten/bitcoin-isnt-so-broken-after-all/, November, 2013.

[17] Bitcoin - Protect your Privacy; http://bitcoin.org/en/protect-your-privacy, November, 2013.

[18] Bitcoin Store ; http://www.bitcoinstore.ch/, November, 2013.

[19] Bitcoin Network Graphs; http://bitcoin.sipa.be/speed-lin.png November, 2013.

[20] Bitcointalk; https://bitcointalk.org/, November, 2013.

[21] Bitcoin Wiki, Block Hashing Algorithm, https://en.bitcoin.it/wiki/Block_hashing_algorithm

[22] Bitcoin Wiki - Blocks; https://en.bitcoin.it/wiki/Blocks November, 2013.

[23] Bitcoin Wiki - Difficulty; https://en.bitcoin.it/wiki/Difficulty November, 2013.

[24] Bitcoin Wiki - History; https://en.bitcoin.it/wiki/History, November, 2013.

[25] Bitcoin Wiki - Mining Hardware Comparison; https://en.bitcoin.it/wiki/Mining_hardware_comparison November, 2013.

[26] Bitcoin Wiki - Pooled Mining; https://en.bitcoin.it/wiki/Pooled_mining November, 2013.

[27] Bitmine; https://bitmine.ch/, November, 2013.

[28] BitStamp; https://de.bitstamp.net/, November, 2013.

[29] Blockchain Info; http://blockchain.info/, November, 2013.

[30] Blockchain Info - Block Size; https://blockchain.info/de/charts/blocks-size November, 2013.

[31] Blockchain Info - Charts; http://blockchain.info/de/charts, November, 2013.

[32] Blockchain Info - Daily Transactions; https://blockchain.info/charts/n-transactions?timespan=60days&showDataPoints=false&daysAverageString=1&show_header=true&scale=1&address= November, 2013.

[33] Blockchain Info - Hashrate Distribution; https://blockchain.info/de/pools November, 2013.

[34] Block Explorer; http://blockexplorer.com/ November, 2013.

[35] Butterfly Labs - Bitcoin Mining Hardware 5 GH per s Bitcoin Miner; https://products.butterflylabs.com/homepage/5-gh-s-bitcoin-miner.html, November, 2013.

[36] CoinCraft Desk (February week 1 batch nummer 2) - BITMINE; http://bitmine.ch/?product=coincraft-desk-january-batch, November, 2013.

[37] Coindesk; http://www.coindesk.com/, November, 2013.

[38] Coindesk - Is Bitcoin legal?; http://www.coindesk.com/information/is-bitcoin-legal/, November, 2013.

[39] Coindesk - What Exactly Does the US Government Really Think of Bitcoin?; http://www.coindesk.com/bitcoin-the-regulatory-story/, November, 2013.

[40] Coindesk - Winkelvoss prediction; http://www.coindesk.com/winklevoss-twins-bitcoin-hit-market-cap-400bn/, November, 2013.

[41] Forbes - Silk Road 2.0 Launches; http://www.forbes.com/sites/andygreenberg/2013/11/06/silk-road-2-0-launches-promising-a-resurrected-black-market-for-the-da November, 2013.

[42] Krebs on Security - Feds arrest Silk Road seller; http://krebsonsecurity.com/2013/10/feds-arrest-alleged-top-silk-road-drug-seller/, November, 2013.

[43] Krebs on Security - Feds take down Silk Road; feds-take-down-online-fraud-bazaar-silk-road-arrest-alleged-mastermind/, November, 2013.

[44] MtGox; https://www.mtgox.com/, November, 2013.

[45] NZZ - Bundesrat nimmt Bitcoin unter die Lupe: Online-Waehrungen; http://www.nzz.ch/aktuell/schweiz/bundesrat-bitcoin-pruefung-postulat-1.18173118, November, 2013.

[46] Reddit- BTC Guild mines six blocks in a row. Again; http://www.reddit.com/r/Bitcoin/comments/1le3rq/btc_guild_mines_six_blocks_in_a_row_again/, November, 2013.

[47] Tazzine ; http://tazzine.ch/, November, 2013.

[48] The Economist - Let them remit; http://www.economist.com/news/middle-east-and-africa/21581995-western-worries-about-money-laundering-are-threat November, 2013.

[49] The Verge - US seizes and freezes funds at biggest Bitcoin exchange; http://www.theverge.com/2013/5/15/4332698/dwolla-payments-mtgox-halted-by-homeland-security-seizure-warrant, November, 2013.

[50] Threat Level - Feds Arrest Alleged 'Dread Pirate Roberts', the Brain Behind the Silk Road Drug Site; http://www.wired.com/threatlevel/2013/10/silk-road-raided/, November, 2013.

[51] Threat Level - How the Feds Took Down the Silk Road Drug Wonderland ; http://www.wired.com/threatlevel/2013/11/silk-road/, November, 2013.

[52] Wikipedia entry for Money; http://en.wikipedia.org/w/index.php?title=Money, November, 2013.

[53] Wikipedia entry for Silk Road (Marketplace); http://en.wikipedia.org/w/index.php?title=Silk_Road_(marketplace), November, 2013.

[54] Wikipedia entry for Tor; http://en.wikipedia.org/w/index.php?title=Tor_(anonymity_network), November, 2013.

[55] Youtube - Bitcoin in the Future Panel - Bitcoin 2013 Conference; `http://www.youtube.com/watch?v=_qdr_Z3hrqQ&feature=youtube_gdata_player`, November, 2013.

[56] Zerocoin; `http://zerocoin.org/`, November, 2013.

Chapter 6

A Success and Failure Factor Study of Peer-to-Peer File Sharing Systems

Christian Lüthold, Marc Weber

With the appearance of the P2P-based music sharing platform Napster in 1999, a new era of file sharing activities was heralded. The service's rapid flourish and ability to accumulate a huge community on the one hand and its prompt fall out on the other, unearth questions about the reasons of its overwhelming success. This work focuses on the responses to such questions by analyzing and comparing six different P2P file sharing systems that are commonly perceived to be successful. As a result of the study, different factors, by which various systems can be compared, are extracted and classified. The investigation and analogy of those factors allow to draw conclusions about miscellaneous system properties, also with respect to time. Therefore, this study not only gives insight into the most common systems and a general comparison, but also enlists core factors, together with their associative relationships and dependencies, and finally enables the compiling of essential properties that might indeed foster file sharing services to gain broad adoption and popularity.

Contents

6.1 Introduction

6.1.1 Motivation

Since the emerge of Napster in 1999, file sharing is well known to the online community. What started as a simple way to share music files between friends, quickly became the most important free source for all kind of digital content to the users. The topic of file sharing is highly controversial. On one hand it allows to easily find, download and share digital content with others. On the other hand it opens the door and even supports the user in committing copyright violations. Besides the booming of content sharing, the peer-to-peer (P2P) technology improved contemporaneously and had a high impact on the sharing community. What was hyped in the beginnings became bad reputation over the years, mainly because of increasing amounts of lawsuits and new laws about protecting the original content providers, like musicians and movie makers. However, in one way or another, file sharing community and technology survived over the last decade. This report investigates the different factors which may lead to the success or failure of a file sharing system. Although the main focus is put towards P2P-based systems, classic client-server solutions are covered to some degree in order to give the full picture. After defining what is understood by the term "success" of a P2P file sharing system, various popular approaches are explored with respect to history, system architecture and other specific properties. This work further compiles a list of factors that may indeed predict the success or failure of potentially new systems. By classifying those factors, individual conclusions can be drawn and assigned to each category. Furthermore, this work emphasizes the impact of the time and describes its part of influencing the chances of a system at any point in time. The investigation and study of the separate factors and the derived conclusions unfold some indications about features and desirable properties that allow file sharing services to become popular and generally successful.

6.1.2 Definition of Success

The first question arising in this paper is how success for a P2P-based file-sharing system could be defined. It is inevitable to mention that there is not one single view, but there are various different perspectives for which success can be specified. As the most intuitive dimension, the user base could be taken into account by simply stating that the more participants the platform has, the more successful it is. On the other hand, the system developers might have additional criteria, such as technical details, that even allow the system to get popular based on intrinsic features that are hidden from the application perspective. Also, success might be to generate money out of the application, either by selling the software or services, or getting funded in a completely other way. Another view concerns the content industries, where the intention is clearly to fight against the emerged piracy issues. Thus, developers need to find ways to avoid technical and legal attacks since the industry is trying to bring down each and every possible source of illegal content distribution. Furthermore, there are maintainers and donors whose aim is to support the system with work power and content, respectively. In exchange they might like to be rewarded, often just with reputation and fame, but sometimes also with real currency. However, as mentioned above, users build the most important group of stakeholders as they provide the content and most of the needed hardware. But users have a lot of different interests and desires and hence their requirements are not homogeneous at all. A sharing platform raises and falls through the participation of users by design, even more in the world of P2P technology. Therefore, the services need to satisfy the wishes of their clients as closely as possible. Unfortunately, such user requirements may end up diametrical or even conflicting with system requirements. This report will show that some sort of trade-

off is unavoidable. Also, taking into account more economical factors, a system needs to survive to some degree. Therefore, durability is a clear success factor, if not even the most important. This requires to cope with ongoing change, such as upcoming and more lucrative competitors and fluctuation in network size. A system's dimension, however, is a coin with two sides: first, it is good for the system because more content and resources are available, but at the same time keenly provokes possible opponents to attack the most popular systems first. Deductively, success is both survival — meaning the protection against any harm —, as well as fulfilling the needs of the system's users.

6.1.3 Outline

Aside the definition of success and an overview of related studies in Chapter 2, this work's purpose is threefold. First, various file sharing approaches that are supposed to have achieved major successes are explained in detail. In favor of this, each system's history and architecture are presented. Furthermore, besides general advantages and disadvantages of each approach, it's economical factors are listed and the most essential properties are identified. All this is embraced by Chapter 3. Second, the extraction and weighting of individual potential success factors as well as their categorization, discussion and evaluation are done in Chapter 4. Ensuing and third, Chapter 5 provides a short recapitulation and explains potential success reasons. In addition, this chapter proposes some desirable properties for futuristic file sharing systems based on the most significant derived factors. Finally, Chapter 6 summarizes and concludes by reflecting the topic and proposes some directions for further research.

6.2 Related Work

To define and prove success or failure of peer-to-peer file sharing networks is hard to accomplish by formal means. Nonetheless, various researchers have analyzed and shown individual factors or features that are of essential importance or may have a negative impact respectively.

Hu et al. [8] specialized on factors that affect individuals usage of peer-to-peer sharing in order to give the investigation of this relatively new technology and its various uncertainties a start. Those uncertainties, such as resource piracy, computer attacks by malicious peers and privacy invasion are assumed to directly influence the usage. They state that users could be exposed to uncertainties related to the peers (e.g., malicious interests), the vendor or creator of the P2P sharing software (e.g., disclosure of online activity) or the Internet (e.g., man-in-the-middle attack). Base on trust literature, the researchers developed a trust-risk-intention model and identified several trust antecedents in the P2P sharing context that might be potentially useful to creators in order to shape and justify their decisions. Hu and his colleagues argue that without proper control of the risk in P2P sharing, users may choose not to use the software due to high risks and thus reduce its success. Further, their work proposes three facets of the perceived risk, namely performance-, privacy- and legal risk. They conclude, that user's intentions to reuse P2P-based sharing software depend on both trust in the software vendor and risk perception associated with P2P sharing. Vendor's thus should actively take efforts in addressing issues like free-riding, content piracy, malicious computer attacks and provide peers with incentives.

Lee [5] is also shedding some light on the user perception of P2P systems by reporting what is important for them to make such systems successful. Instead of debating from technical or legal views, he focuses only on the less explored end-user perspective which is critical for designing better technology. Lee claims that system builders need to know

what their potential customers consider most important in P2P-based file sharing systems (e.g., how much they tolerate free-riding). By means of surveys and observations, his work compiles a list of representative P2P sharing features and extracts groups and relations among those.
Other propositions are more directly targeted at the P2P system architecture of distributed sharing software. So is the work of Huang et al. [4] which criticizes the poorly scaling flooding-based query algorithms used in many overlay networks, and propose novel solutions to improve the search performance in distributed file sharing systems. Furthermore, they provide several guidelines and techniques for constructing so-called desirable topologies and conduct extensive experiments to justify the performance gains.
Lots of investigations on basic P2P properties have been made by various academic works. Bhagwan et al. [1] investigate on the term "availability" and try to define its understanding for peer-to-peer systems. They point out that the principal challenge in designing highly available systems is the toleration and recovering of failures. Network designers thus are required to anticipate transient software, as well as possible hardware, failures, partial or total communication interruption and the ongoing fluctuation of participating and leaving nodes. The team further examines several characteristics of host availability and discusses the implications of their findings on the design of P2P systems.

6.3 Analysis of Existing Systems

In this section we are going to explain the technical details of existing systems. Our aim is to give a good overview about the used architectures and protocols. Furthermore we analyze the economical factors of each system and then conclude each section with a list of positive and negative aspects of the analyzed systems and a short conclusion. In order to do so, we start with the earliest concept and dig further to the more recent technologies.

6.3.1 Early Napster

6.3.1.1 History

Whenever the topic deals with file sharing, the original Napster drops in as prime example. As it was the first system whose main purpose was the sharing of digital content, this paper treats Shawn Fanning's project, which started in 1999, first of all, too. Napster was envisioned as an independent P2P file sharing service and primarily focused on the exchange of music files and made use of the relatively new MP3 format which allowed digital music files to be small enough - compared to the back then established WAV music disk format - to be sent over the web. Although the system initially was just intended to serve the creator's own needs, it rapidly grew in terms of users willing to share their music libraries. According to [10], the true genius of Napster was not that it let users trade music over the Internet, but that it made peer-to-peer networking so simple and easy-to-use. Napster took both the mp3 format and the concept to the masses. However, due to the ongoing copyright infringements, the Recording Industry Association of America (RIAA) filed a lawsuit against the service which lead to an injunction and finally to the shutdown of the service in 2001. Napster had to agree to pay a settlement to music creators and copyright owners for its unauthorized uses. In order to do so, Napster attempted to convert their so far free service to a subscription system which emerged and developed until today.

6.3.1.2 Architecture

The Napster system architecture, depicted in Figure 6.1, focused intensively on the usage of a central server and thus was not a pure peer-to-peer system but rather a hybrid one. All client nodes, called peers, directly contact this central server in order interact with the network. This made the bootstrapping problem simple, because clients just registered on the central instance and subsequently were allowed to query the central index. Whenever clients connected to the server, they uploaded a list of music file they were willing to shared. Based on this incoming information, the server managed and updated the globally central index while optimizing it for quick and efficient searches. Such searches had the simple form of clients asking and the server returning a list of possible hits. Ensuing file downloads, or rather transformations, were possible by establishing a direct link from the requesting client to the serving peer.

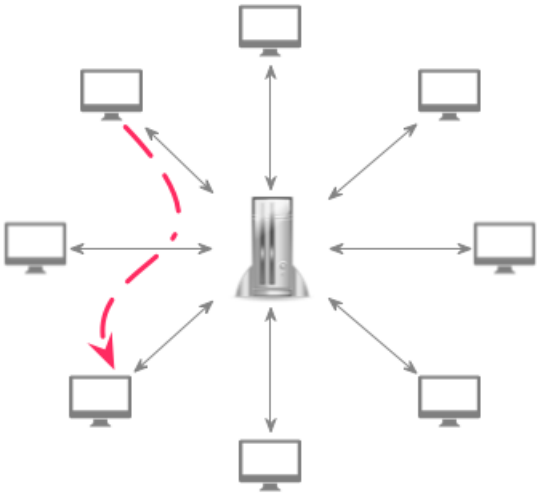

Figure 6.1: The Napster architecture.

6.3.1.3 Economical Factors

Economical interest in the "age" of Napster was mainly on the side of the music industry (they had the feeling that they lose some money) and on the client side (music for free). The creators and maintainers of the system did not have any financial interests and also no business model.

6.3.1.4 Pros and Cons

The most obvious element in the Napster architecture is the centralization of the single main server. This approach assures an easy bootstrapping mechanism as well as user registration and authentication. Furthermore, a centralized index guarantees a complete and fast search and at the same time minimizes the chances for data inconsistencies and outdated search results. The good search facilities are combined with a P2P-based exchange facility of the actual content, which is music files. So basically, Napster combines the best parts of two worlds: user and content management over a centralized entity, as well as distribution and storage of the content on the client side over a P2P system. One of the major drawbacks of this approach, in the context of file sharing, is its vulnerability, namely the exposure of the central unit. This difficulty is often referenced as "single-point-of-failure" since it is easy to shutdown the whole system by simply attacking such a unique and central element. Such attacks are possible both on the technical (e.g., DDoS) or legal (e.g., court order) level. The existence of copyright infringed material can easily

be proven due to the early Napster's architecture that allows to install the free client and query the database for suspicions. Another drawback concerns the file speed transmission. A downloading client is limited to one file source and thus depends on the serving peers upload speed.

6.3.1.5 Essential Properties

To summarize the main attributes of the Napster service it is important to focus on its central indexing server, that eased both the bootstrapping problem and the user management, like registration and authentication. But its main purpose was the management of the overall music file index. Due to its single instance, any searches could be optimized in terms of speed and completeness, where even fuzzy queries could be answered. The early Napster thus combined the best parts of two worlds: the user and content management system that is very known from client-server solution and the distribution and content storage mechanisms as peer-to-peer approaches. It is also worth mentioning that the service was the first of its type. The free and easy-to-use client satisfied the users completely in this relatively early stage of file sharing. At the time, sharing simple and small music files sufficed. However, the central entity turned out to be a "single-point-of-failure" because it was vulnerable to both technical and legal attacks. In addition to that, the existence of copyright infringed material could easily be proven by just querying it, what made the collection of evidences very simple.

6.3.2 Gnutella

6.3.2.1 History

When talking about file sharing applications that rely on peer-to-peer technology, another important part in history acts Gnutella. To date, although its name references the software client, it mostly refers to the underlying network protocol. The inventor of Gnutella, Justin Frankel, was inspired by the Napster idea but envisioned concrete improvements over Fanning's system. On the one hand, the Gnutella protocol should not depend on any central entities in order to avoid juristic accusations. Further, the end clients should communicate directly with each other so as to prevent any centralized service bottleneck by design. Another improvement would allow users not only to share music, but files of any kind. The project started at Nullsoft, a company founded by Frankel and later bought up by AOL, with the intention to release the client as closed-source. However, AOL prohibited its distribution the same day it was published in 2000. It is due to some external programmers, who achieved to reverse engineer the client, that the Gnutella distribution continued. While the software lives on as an open-source project until today, it is not part of the GNU project and its name rather arose from Frankel's aim to license the application under the GNU General Public License. Gnutella had a high impact on the peer-to-peer file sharing community, not only because the service was not restricted to music files only, but also because the underlying protocol allowed diverse client applications to develop and fit various needs. The most popular Gnutella client, LimeWire, registered up to 28 million users at its all-time high in 2007. In 2010, however, LimeWire was forced to shut down and caused a notable drop in the network size [9].

6.3.2.2 Architecture

The Gnutella architecture is inspired by the friends-of-friends principle where requests are forwarded from any node to its friend nodes. With this approach, the amount of inquired nodes grows exponentially with each forward. In order to bootstrap to the network, Gnutella run through several techniques until today. Pong-Caching represented

the initial approach where nodes exchanged their respective bootstrap addresses such that the probability of finding a starting node in the next session was optimized. However, such caching did not solve the problem for absolutely new nodes that had not a single contact when joining the first time. In order to address this issue, Gnutella introduced the GWebCaches in a more recent update. By means of such caches, the very first bootstrap can be done by contacting a node that is openly put out to tender. Once a user successfully connected to the network, file searches can be queried. This is implemented by sending the query to at least one known contact, which forwards it to its contacts in case he does not own the file which was asked for. Here again, the friends-of-friends principle ensures a certain spreading factor which increases the investigated space exponentially. This approach has the big advantage, that it works all the time, as long as the client gets at least one other machine running the Gnutella software. However, two issues had to be addressed by the developers of the protocol: avoiding circles and limiting the search. Nodes that already forwarded an incoming query before, do not forward it again. Furthermore, a so-called "time-to-live" (TTL) parameter is passed along with the query and decreased by every passing on so as to stop the query at a certain level of depth. Once a node who can serve the result is reached, it directly contacts the requester and delivers the queried content. This is basically how the content searching works in Gnutella, however, client mostly do not have enough resources (i.e., CPU, bandwidth) to handle thousands of requests every minute. The protocol therefore introduced the concept of ultra peers and leaves, which can both be seen in Figure 6.2. Ultra peers represent nodes that are specified to maintain address books, namely the knowledge of which leaves contain what files, and indexes only, whereas leaf nodes are the actual content holders. Requests thus are sent to the nearest ultra peer, which routes the request to a friend ultra peer and so on until one ultra is found who knows a leaf node that stores the searched content. With this technique, only ultra peers that possess enough resource capabilities are involved with the routing of searches, whereas the leaves concentrate on storing and providing content. To further improve the search, the "Query Routing Protocol" (QRP) was introduced which allowed ultras to communicate with each other and exchange hashes of the files they have the direct address to. With this protocol, the routing tree of a specific query can be pruned in such a way, that the message is forwarded only into the direction of the address holder. Another improvement to lower the network traffic was to use "dynamic querying" where ultra peers forward messages only to one further ultra peer at the time and wait for a satisfying answer before continuing. Usually, the Gnutella protocol not only returns a single target address that stores the requested file, but up to 250 contacts. This makes sense regarding the fact that not all contacts may be available or have enough resources at the time to serve. Whenever a requesting node wants to receive content from its host, it informs the contacted host about the other hosts that were found with the query. By doing this constantly, the Gnutella protocol establishes a so-called "Download-Mesh" that allows leave nodes to know about other hosts of a specific file such that they can forward download request in case they are too busy to handle them themselves. Furthermore, this approach supports "swarming", a concept that allows to download a file from different hosts simultaneously where different parts of the whole file are loaded from different machines. This allows receivers to become servers of file parts they have already loaded and are able to contribute to the Gnutella network much more quickly than waiting for the whole file to complete. In order to avoid the download of corrupted parts, so-called "file-magnets" were introduced. Those meta file basically are pre-configured enlistments of hosts that are able to serve parts of the files and further provide a way to proof that the correct file was loaded, by comparing the arrived file parts' hashes with the configured ones.

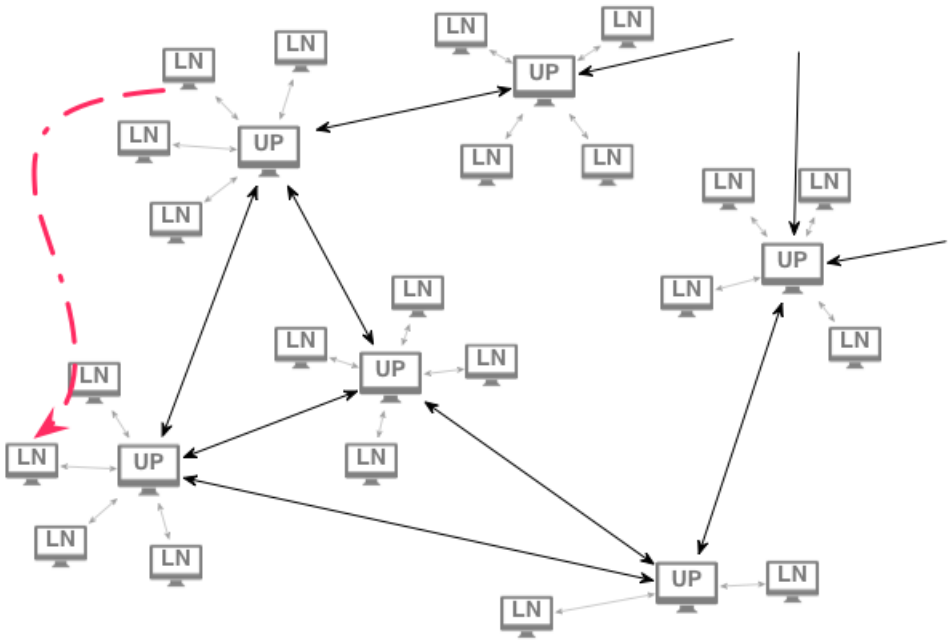

Figure 6.2: The Gnutella architecture.

6.3.2.3 Economical Factors

When Justin Frankel developed the Gnutella network protocol his main driving force was never to drain money from the result although it is worth to notice that he was already worth more than 100 million dollars, due to the buyout of Nullsoft and the voluntary shareware fees of his former Winamp project, at that time. The inventor knew that he would never get any money from it and that such technology would rather be good karma as it empowers people. It is safe to assume that those reasons drove the whole Gnutella community to continue and improve their services since they were aware of the fact that they enabled new ways of digital consumption for themselves. However, when concentrating on the diversity of Gnutella clients, some can be found that actually achieved to make money. The very popular LimeWire client had several sources of revenue before it was ordered to shut down. First, the application-integrated store offered a way to purchase music legitimately. Second, a premium version of LimeWire was sold that offered various improvements, like searching more connections and thus increasing download speed, over the free version. Third, it included bundled software as part of the client installation, getting revenue from external sources which included toolbars or similar as part of the default installation.

6.3.2.4 Pros and Cons

Overall, the Gnutella network protocol has many advantages over Napster. For example, the system scales much more nicely than Napster's due to the pure peer-to-peer technology. Further advantages of its implementations allow queries to be possible at any time, no matter how much of the network is currently down. Also, the way requests are forwarded and handled allows to explore an enormous network space where also fuzzy searches return results. From a user's perspective, Gnutella is an attractive file sharing approach because it is simple to use, offers a broad variety of user-friendly clients and supports the sharing of files of any type and size. Furthermore, clients are always backwards compatible since they are implemented with respect to the GDF specification of the Gnutella community. On the other hand, the protocol also shows some drawbacks. Although quite some effort has been put into network load optimization, the friends-of-friends principle, often also

referred to as "flooding" the network, cause the network to scale badly. In addition, even though thousands of computers may be reached, it is never sure that any of them contains the files the search is looking for. Also, the search investigations are not instantaneous but consume quite some time as they have to bounce around so many nodes. Another downside is the yielding of resource sacrifices a machine has to obey as soon as it is part of the Gnutella network. Otherwise, the decentralized manner the peer-to-peer network is operating with would not work because such resources have to be given up in order to handle incoming requests from other users. Even though the protocol works in a fully decentralized manner, there are still some weak points where potential attacks could decrease the service quality. Such activities could have the form of provisioning massive amounts of junk files or flooding the network with unnecessary queries. Furthermore, external services that rely on centralized information might pull the attention of attackers, too. Examples could be websites that host Gnutella file-magnets or client software itself.

6.3.2.5 Essential Properties

The open-sourced Gnutella protocol allowed many software clients to emerge, each of them being backwards-compatible to its predecessor (due to the community's GDF specification) and having own specific features (e.g., BearShare comes with encryption). In comparison to Napster, Gnutella also allows any kind of file format to be shared. Further, the focus laid on implementing the system in a fully decentralized manner. On the one hand to reach a better scalability that allows the network to run even if large parts of it are down, but also to avoid legal issues. Interestingly enough, the industry seems to attacks rather popular or trademarked clients (i.e., LimeWire) instead of the protocol, as it is untouchable. But although single clients get prohibited or stay under a court order, new clients emerge and distribute quickly since they are able to operate on the exact same network as the client before. However, the "flooding" approach used in many technical scenarios turned out to be the major drawback of the protocol and thus experienced many optimizations over the recent years.

6.3.3 KaZaA

6.3.3.1 History

KaZaA uses FastTrack, a proprietary protocol. Both KaZaA and its protocol were created by Estonian programmers from BlueMoon Interactive and sold to Niklas Zennström and Janus Friis. The KaZaA Client called "Kazaa Media Desktop KMD" was introduced in March 2001 by the Dutch company Consumer Empowerment. From the beginning KMD was bundled with advertisement and malware. Therefore about a year later, the first unofficial modification of KMD became available. These clients with names like "KaZaA-Lite" were created by a group of third party developers and downloadable free of charge. The benefit of these "hacked" clients were no advertisement and a faster search function. Right from its start on, KaZaA was under legal attacks. In 2001, legal action in the Netherlands forced an offshoring of the company to Australia, and a renaming to Sherman Networks. In September 2005 Sherman Networks was forced by Australian court, to modify the client software in such a way, that Australian citizens were not able to share copyrighted content. The company refused to do so but instead showed a banner on the website, stating that the software may not be used by Australian users anymore. At the end of July 2006 it was announced that Sherman Networks finally settled with the record industry and motion pictures studios. The company agreed to pay $100 million to the four major music companies and to equip the client with a filter for copyrighted content. In 2007 a new version of KMD became available without spy-ware and malware but still containing the toolbar and advertisements. In 2009 KaZaA changed its business

model one last time to the form of legally redistributing content for a monthly flat rate of $19.98. The distributed content however was DRM protected. Finally in 2012 KaZaA ended its services and closed down.

6.3.3.2 Architecture

Although the protocol (FastTrack) behind KaZaA is proprietary and therefore closed source, most of its internals are known today. Based on [6] we know that the protocol builds a pure peer-to-peer network with no central element. However the KaZaA networks takes into account, that not all peers are equally powerful regarding availability, bandwidth connectivity, CPU power, and NATed access. It exploits this heterogeneity by organizing the Peers in a two-tier hierarchy consisting of Super Nodes (SNs) in the upper tier and Ordinary Nodes (ONs) in the lower tier (see Figure6.3). ONs are connected to their SN through semi-permanent TCP connections. The SN again maintain TCP connections to other SNs creating the top level overlay of the network as shown in Figure 6.3. The SNs are sparsely connected, each SN is connected to about 0.1% of the other SNs. Super nodes also act as proxies for ON which are behind a NAT or have dynamic or random ports. Each ON may potentially be promoted to become a SN. Still today it is not absolutely clear, how the promotion works. Some of the essential properties needed for a promotion are: Not being behind a NAT and in general have more bandwidth and CPU power than ONs around your node. All communication between the peers is encrypted. The protocol uses four different types of TCP traffic:

1. signaling traffic used for connection establishment and meta data exchange
2. file transfer traffic for the actual file download
3. commercial advertisement sent over HTTP
4. instant messaging traffic

Another characteristic of the FastTrack protocol is that in both tiers, connections are highly dynamic. On average the duration of an ON to SN connection is 34 minutes. Tier one connections are even shorter: A SN holds a connection to another SN for 11 minutes on average. Further more, about one third of all ON to SN and SN to SN connections last less than 30 seconds. There may be two reasons, which may explain these short durations:

1. During the bootstrapping process one ON connects to several (normally 5) SNs. Once the ON has selected its parent SN it will disconnect from the other SNs
2. Third party clients like KaZaA-Lite "hop" between several SN during a search, which causes a lot of disconnects and reconnects.

Information about the available SNs are periodically exchanged between SNs as well as between SN and ON. This information is physically stored on each node and used during the bootstrapping process. Each ON holds a local list of SN addresses. In the bootstrapping process, some of the cached SN are selected (typically around 5) based on locality and workload. The ON then tries to establish connections to these SNs: First UDP packages are used to probe which SN are available. To all responding SNs a TCP connection is established and bootstrapping data is exchanged: The ON uploads meta-data about the files it shares to the SN and in exchange the SN pushes an list of SNs(IP, port and load) back to update the list on the ON. Next the ON selects one parent SN, based on the locality and load and disconnects from all other SNs. Content management is done in a two step process which again reflects the two tier architecture: For each file a node is sharing, the following meta-data are published to the network(ON to SN): The

file name, the file size, the Content Hash and the file descriptor. Each SN is responsible to create and maintain an index of available files of his ONs. The index maps the file identifier to the IP address and port of the providing node. However these indexes are not shared among the SNs. Basically each SN and its ONs can be seen as a miniature version of a Napster network with the difference, that the SN is not a Server but an ordinary client machine which fulfills a special role in the protocol. The content hash is of major importance during a download request. It is generated by the KaZaA client by hashing the file and thereby producing this unique identifier. In a download request only the content hash is used. Therefore if the download from a specific peer fails, the requesting peer can automatically search for the desired file without having to issue a new keyword query. Because there is no central or shared index, a search query needs to be propagated through the network. However the FastTrack protocol does not initiate flooding searches like the Gnutella protocol but instead uses the fact that the network connections are highly dynamic: Search is done by an ON sending a query (keywords) to the SN. For each match the SN returns the meta-data of the file as well as the IP address and port number of the providing peer. If no match can be found on the current SN the query may be forwarded to some other SNs. Interesting here is the fact, that a query is not forwarded to many SNs thereby a certain search will initially cover only a small part of the overall network. But because the network is highly dynamic, time will solve this problem: The ON re-sends unanswered queries after some time. Because both, the current SN, as well as its other SN connections may have changed in the meantime, the query will reach other parts of the network and may be answered. So instead of flooding the network with queries, the time to answer a certain query is increased. Another, faster approach is used by most third party clients like KaZaA-Lite: The ON sends the query to its parent SN. Once he gets the results back (may be empty) it disconnects and reconnects to another SN in the network. After several hops the search terminates.

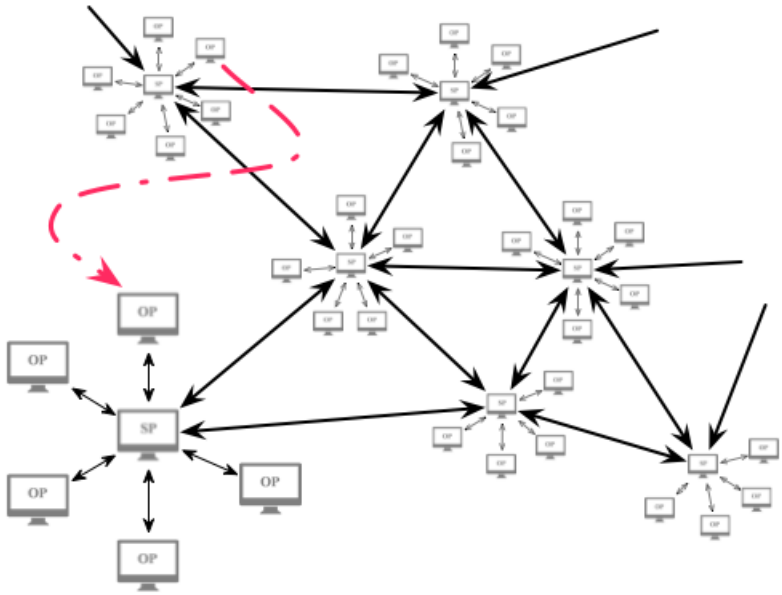

Figure 6.3: The FastTrack architecture.

6.3.3.3 Economical Factors

Right from the begin the idea of generating money was included in the design of FastTrack and KaZaA. One of the four TCP traffic types used by the peers is reserved for commercial advertisement. The KMD had a constantly visible banner presenting advertisements to the user, opened pop-ups of third parties and was bundled with malware. The simple fact

that Sherman Networks could accept and pay a $100 million punishment shows that this business model seemed to work. Although third party clients bypassed the advertisements and corrupted the search mechanism the system was robust in both the architectural structure as well as the revenue channels to withstand this free riding parasites. The final business model with monthly flat rates worked for some years but finally could not prevent the shutdown of the service.

6.3.3.4 Pros and Cons

FastTrack is based on a fully decentralized design, which takes the heterogeneity of the different peers concerning their availability, bandwidth, NATed access and CPU power into account. The two tier architecture is reducing the communication overhead because only the SN connections are building up the overall network. Information about locality and load is used to balance the network. The intentionally build in dynamic of the network (connection shuffling) is used for the restructuring of the network, enables the search and produces a high churn rate which is in contrast to other peer-to-peer systems beneficial for the network. The stronger SNs help overcome problems like dynamic and random ports and provide NAT circumvention if needed by acting as a proxy to their ONs. The build in advertising, spy-ware and malware guarantees the operators a constant source of income but at the same time annoys the users. Therefore third party clients emerged which parasite the system with a free riding behavior. The search does not use a flooding approach which reduces the amount of messages exchanged by the SNs and therefore makes the network more scalable. On the other hand this implementation trades scalability and network message reduction for search speed and completeness. A search with KMD may take a considerable amount of time, especially for unpopular content and there is no guarantee to find the content at all even if it is offered from some nodes in the network.

6.3.3.5 Essential Properties

The FastTrak protocol is proprietary and closed source. The network is fully decentralized. Its two tier architecture of the overlay takes locality and load into account. The intentionally build in high dynamic (connection shuffling) is used to balance the network, enable message cost optimized search and makes the system churn resistant. SNs act as proxies if needed, to overcome the problems connected with firewalls, NAT and dynamic and random ports. Commercial interest are built into the protocol as well as in the official clients. The excessive use of advertising, spy-ware and malware annoyed a lot of users which turned to "clean" third party clients. The network could withstand unofficial clients.

6.3.4 FreeNet

6.3.4.1 History

FreeNet was envisioned by Ian Clarke and first described in his paper "FreeNet: A Distributed Anonymous Information Storage and Retrieval System" [2] in 1999. In contrast to the systems presented so far, FreeNet is designed to fulfill one goal: Protecting the freedom of speech for digital content. This is done by implementing a strong anonymity protection. In 2000 Ian's paper was the most cited computer science paper according to Citeseer. In March 2000 the first version of FreeNet was released which caused a considerable echo in the press all over the world. But the main focus of interest lied on the effects FreeNet might have on copyright content and its illegal distribution and less on

the idea of providing freedom of speech through the Internet. In 2001 Frost [1], a usenet-like messaging system with included download manager functionality became available. In order to improve privacy and make it harder to detect FreeNet nodes, the idea of a Darknet (FreeNet version 0.7) was envisioned, where only trusted peers are used to build the network. The ideas about a Darknet were first presented in 2005, in April 2006 a first alpha version was released, and the final version (v.0,7,5) including Darknet was released in May 2008.More information and downloads of the freenetclient are available on the FreeNet website [2]

6.3.4.2 Architecture

The whole architecture is centered around the idea of a digital tool set which provides complete freedom of speech to its users. The developers claim that true freedom of speech is only possible through complete anonymity, meaning both, the identity of the creator as well as the one of the receiver of content needs to be hidden. Therefore the main design goal behind freenet is to guarantee anonymity to all participants. To achieve this, each node in the network is treated exactly the same, and all content is split up and encrypted. The strict focus on anonymity has a direct consequence to the content management: Content can only be added to the network, but never actively removed! Because the creator (owner) of the content is unknown and therefore not identifiable, users can only insert data, but there is absolutely no way to delete data again. Due to the limited disk space and churn however, unpopular content which was not requested for a long time, will be removed to give room for new and popular content. The network is built up by nodes which randomly connect to other nodes they can find forming a unstructured graph. To participate in the network a node needs to provide three resources:

1. **Disk space** A predefined amount of disk space (adjustable by the user) needs to be reserved. FreeNet divides the space into two caches. A short term one, where all data that the node transfers is stored temporarily until it is pushed out (on a random base) by newer content. A long term one which is designated to contend which the current peer is relatively closed to. Beside this some memory is needed to keep track of other nodes, active request and to check the correctness of data content. Each node maintains a data store containing documents associated with keys, and a routing table associating nodes with records of their performance in retrieving different keys.

2. **Bandwidth** Each peer needs to forward requests and responses of other peers in order to serve other users needs.

3. **Processor power** Each content traversing a peer is checked. Therefore the content key is recalculated and checked. This is done to prevent corrupted data (through transmission or willingly inserted) to be forwarded to other nodes.

Each node has a random identifier which lies between 0 and 1. This identifier is used in the routing process. Nodes which are close according to their identifier tend to store content with lexicographically close keys. Although the identifier of a node is used for routing and to determine where to send requests, a node may, at any given point in time, decide to switch to a new and completely random identifier. This is another mechanism to prevent external parties from tracing down a single node. As mentioned earlier, content is split and encrypted. This is done to protect the individual nodes in cases where their offered storage hold parts of illegal content. In a simple view of the system one can think

[1] http://jtcfrost.sourceforge.net/
[2] https://freenetproject.org

of the key being used to encrypt / decrypt and as well as to retrieve the data it from the network. The real mechanics are a bit more complicated and explained in detail in []. FreeNet offers three types of keys:

1. **keyword-signed key (KSK)** where the key is derived from a simple short descriptive string
2. **content-hash key (CHK)** where the key is created by building a hash of the content
3. **signed-subspace key (SSK)** which creates a name space which can be used to sign content created / inserted by one party while at the same time still guaranteeing anonymity

FreeNet does not provide any kind of a search. The only way to retrieve data from the network is over the content key. Therefore a user needs to know what she is searching for in order to get it. To solve the problem of finding unknown keys for interested content, users may build up special files containing indexes or use clients build on top of FreeNet like for example Frost. To retrieve data a request message is generated containing the

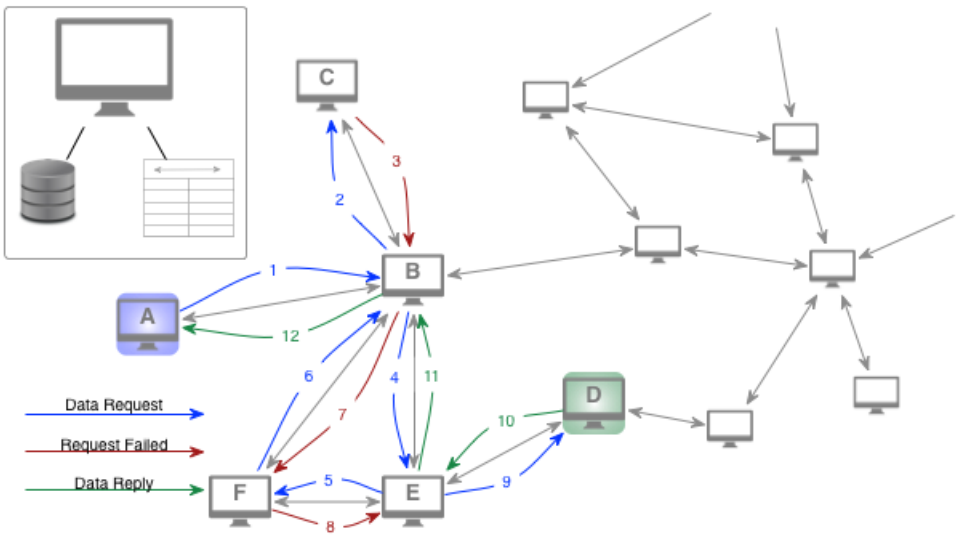

Figure 6.4: The FreeNet architecture.
A typical request sequence. Node A is requesting content which is currently available on node D. Once the request reaches node D (step 9), a copy of the file is stored along the path D-E-B-A.

key of the searched file and a hops-to-live value. This message is then send to the own node to start the search procedure. On receiving a search message, the node first checks if a file for the given key is stored in the cache. If a file is found it is returned together with a note saying it was the source of the data. If not found the nearest key in the routing table is looked up and the request is forwarded to the corresponding node. If the data is found and returned the node checks the validity of the content, passes the data back to its upstream requester, stores a copy of the file in the cache and creates a new entry in the routing table associating the actual data source with the key. To guarantee anonymity a node may choose to alter the data source information. The search works in a steepest-ascent hill-climbing search with backtracking as depicted in Figure 6.4. A node forwards a request to the best node according the key first. If he gets a failure message back it tries the second best node and so forth till it gets the data back or runes out of nodes in which case it will send back a failure message to its upstream requester. Loops

are broken, by sending a failure message back, if the search message already traversed the node. Two things are important: First, a node receiving a search request does not know where the request originated, he only knows the requesting (upstream) node, by which the anonymity of the original requester is guaranteed. Second, if a result (content) is handed back on the reverse search track, each node stores a copy of the content, which makes requested content to be spread over the network and making it more available. Storing content is using almost the same mechanism as requesting it. First a key is generated. Next an insert message consisting of the key and a hops-to-live value is created and sent to the own node. The message is forwarded from node to node like a request message. Each receiving node checks its local cache for the key. In case the key is already taken, a collision occurred and the data for the key is returned just like a request was made. The collision will be detected by the inserting node, which then needs to select a new key and start the procedure again. In case no collision is detected along the whole path, the final node (the one where the hops-to-live counter reaches zero returns an "all clear" result to the sender indicating that the key is available. On retrieving the positive result the inserting node sends the data, which will be propagated along the path created by the request query and stored on each node along the path. As mentioned in the history section, the current version of FreeNet (v.0.7.5) provides a so called "Darknet" feature. This is simply a restriction on which other nodes one node should connect to. While in the "Opennet" mode a node detecting another available FreeNet node simply builds up a connection to this node. In the "Darknet" mode only nodes specified by the user (IP address and port) are chosen to build up connections. This causes a darknet to grow slower but provides a bigger amount of protection because each node only connects to peers it trusts, which in turn are connected to peers they trust. This mode is designed to run FreeNet in areas of the world where users may be afraid of being punished for using such a system. Freenet in darknet mode is almost impossible to detect.

6.3.4.3 Economical Factors

Freenet was built with one major goal: Providing full anonymity to all its users. There were no economical interests involved at any stage of the design, development and operation. The only source of income for the system and its operators is through donation which can be placed over the Freenet Webpage[3].

6.3.4.4 Pros and Cons

Freenet offers a strong privacy protection by providing anonymity to its users. The downside of this protection is the impossibility to provide a search functionality, the identifier (key) of a file needs to be known. The system is mainly focused on the exchange of thoughts in textual form and therefore favors smaller files. Although there is no limit to the size of the content, lager files tend to take longer to be distributed through the network because each node needs to recheck by calculation the correctness of the transported content.

6.3.4.5 Essential Properties

A Freenet node can be run in one of two modes: Opennet or Darknet, which differ in how easy they can be detected by third parties. The architecture is designed to offer maximum possible anonymity to its users and is in this category currently the clear leader. Content is split, encrypted and distributed over the network. There is no search available and possible, content can only be retrieved (and decrypted) by knowing the key. Content can

[3]https://freenetproject.org/donate.html

only be added but not actively removed and "interested" content (many requests) will be automatically redistributed more which leads to more redundancy.

6.3.5 BitTorrent

6.3.5.1 History

The BitTorrent protocol was invented by Bram Cohen in April 2001 for the online commuity of etree[4]. The main focus lied on the fast redistribution of large data files. In July 2001 the first implementation was released. Quickly the file-sharing community became attracted by the technology and started to use it, which boosted the popularity of the system drastically. 2004 BitTorrent Inc.[5] was founded which is responsible for the development of the protocol itself as well as for the two clients "μTorrent" and "BitTorrent Mainline". In 2005 Azureus (today known as Vuze[6]) was released which included a trackerless mode. This client was able trackers by a Distributed Hash Table (DHT). In the same year BitTorrent Inc. released a new version of their client called "Mainline DHT" which offered the exact same functionality. In 2007 BitTorrent Inc. changed their Website to an online store and started to redistribute music and movies. The final version of the BitTorrent protocol was released in 2008.[7] According to [7], "*BitTorrent is the most popular peer-to-peer application*" in 2012.

6.3.5.2 Architecture

In contrast to the systems discussed previously, BitTorrent does not form one big network but individual ones so called "swarms" around each content (file or directory containing files and subdirectories). The content of a swarm is divided into segments called "pieces" and each swarm is managed by a tracker. Since 2005, a trackerless mode is available as well. But the only difference is, that the tracker is replaced by a Distributed Hash Table (DHT). Because the basic mechanisms are the same for modes, this article only explains the mechanism involving a tracker. As depicted in Figure 6.5, a BitTorrent swarm consists of three elements:

1. **.torrent file** holding the information about the tracker(s) needed to bootstrap to the swarm, a short description of the content and a checksum for each piece of the content for validation.

2. **Tracker (T)** (optional since 2005) which coordinates the peers in the swarm by keeping track which peer offers which pieces of the content and distributing this information to the peers.

3. **Peers (S and L)** which are actually sharing the content. Peers fall into one of two categories: seeders (S) which hold a full copy of the content and only upload pieces to other clients and lechers (L) which have parts of the file and are simultaneously up- and downloading pieces of the content.

In order to download a certain copy the user first needs to get the corresponding .torrent file. This step has to be done individually (e.g., exchange with friends, download from a webpage etc.) and is not part nor supported of the BitTorrent network. The .torrent file is interpreted by the client which first contacts the tracker and informs the tracker about the pieces the client can offer (at the begin empty). The tracker responds by sending back

[4]http://etree.org/
[5]http://www.bittorrent.com/
[6]http://www.vuze.com/
[7]http://www.bittorrent.org/beps/bep_0003.html

a list containing the addresses of some clients known to him, together with the pieces they offer. The client then starts to ask other clients of the swarm for a certain piece of the content. Once a piece is downloaded correctly (checked using the checksum) the tracker is informed about the new state (available pieces) of the client, and receives an updated list of other clients. This process is repeated till all peaces of the contend are downloaded. A this point the client becomes a seeder and only uploads pieces to other clients. To enforce fairness in the system a "tit-for-tat" approach is used: The requests of peers which upload content are preferred to those of peers who only download. This is also the reason why BitTorrent downloads start slowly but speed up over time. The BitTorrent protocol operates over TCP and /*mu*Tp (The motivation for /*mu*TP is for BitTorrent clients to not disrupt internet connections, while still utilizing the unused bandwidth fully.[8]) Because peers connect to each other over TCP to download the pieces, clients sitting behind a NAT can't be reached. Basically BitTorrent (the protocol) is not doing anything in order to overcome the problem, however most clients try to make the users life as easy as possible (UPnP or NAT-PMP) and have tools incorporated to help setting up the network connections properly. A very good example is uTorrent[9].

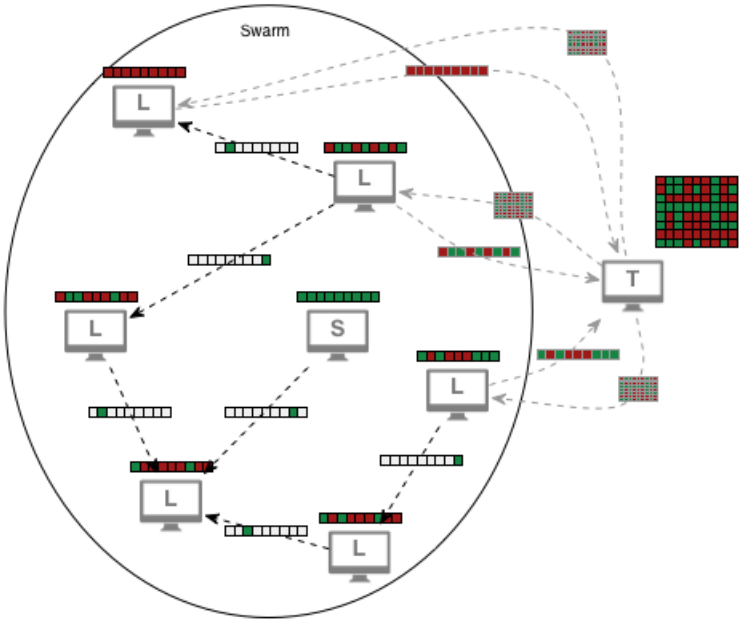

Figure 6.5: The BitTorrent architecture.

6.3.5.3 Economical Factors

The BitTorrent protocol was designed to facilitate the exchange of large content Its potential was realized not just by the file-sharing community but also by industry. BitTorrent Inc. was never under any type of legal pressure but actually became a business partner of the music and film industry by acting a s reseller of their content. Beside the partnerships, the BitTorrent protocol is also used by big companies like Blizzard, Facebook and Twitter, to update their internal server farms for example. This provides another income stream from support contracts.

[8]http://www.bittorrent.org/beps/bep_0029.html
[9]http://www.utorrent.com/intl/de/help/guides/connection-setup

6.3.5.4 Pros and Cons

BitTorrent is currently the most used and best-suited protocol to share large sized content over a distributed network. The simple protocol scales very well and is in its current version sable since over five years. A fair sharing is guaranteed over the included "tit-for-tat" mechanism and because the whole available upload bandwidth of all clients is used, faster and parallel distribution of the content is possible. On the downside we have the lack of a search function, which is completely impossible because the networks (swarms) are build around a certain content but there is no overall network. Another drawback may be the need to configure the own network to allow direct TCP connections. And finally anonymity is not taken into account at all as all addresses of the clients participating in a swarm are exchanged openly.

6.3.5.5 Essential Properties

The BitTorrent protocol facilitates fast and load balanced distribution of large content over a decentralized network. The sharing is done in a "fair" way using "tit-for-tat" and the available upload capacity of all clients of a swarm is used if needed. The lack of a search did not hinder the popularity of the protocol, but on the other hand protected the developers and operators from legal threads. BitTorrent Inc. is a strong company having both, business partners like the content industry and other business as customers (Blizzard, Facebook, etc.) which use the technology in-house or to provide faster and better services to their own clients. The stable protocol is supported by more than twenty different clients and therefore sharing over BitTorrent is possible from all well known operation systems today.

6.3.6 Wuala

6.3.6.1 History

Around 2005, a small team around Dominik Grolimund at the Swiss Federal Institute of Technology (ETH) started to develop a new file sharing system as they faced the time where cloud storage became popular. To be contrary to the commonly used client-server-based online storages, the team rather focused on implementing a distributed file system based on peer-to-peer technology. After three years of development and research, a first open beta of the project, named Wuala and created under the Caleido AG, founded by Grolimund, was available in 2008. Not only was the mission of the Wuala service to store, share and publish files on the internet, it also consequently aimed at making such operations secure and private. Whereas older peer-to-peer file sharing clients allowed to send files over the network as-is, Wuala distinguished itself from other services by the support of client-side encryption of files that entered the network. A year later, in 2009, the French computer hardware company LaCie, especially known for storage systems, announced the acquisition of Wuala. This was a major step in the history of Wuala since the underlying system architecture had to be changed. LaCie demanded that large parts of the peer-to-peer storage technology had to be replaced with the more established client-server solution. Since then, the encrypted user data is not only stored in the peer-to-peer space, but also on servers located in central Europe. Although Wuala never received as much publicity as former peer-to-peer technologies, its closed-source development is highly promoted and improved till this day.

6.3.6.2 Architecture

Wuala makes use of a structured DHT network, the third generation peer-to-peer technology, which ensure a uniform distribution of the content among all the participating peers.

In addition to this structured overlay, a dedicated server is provided to simplify the bootstrapping. The Wuala network defines three node, also depicted in Figure 6.6, classes that each have other responsibilities. Super nodes are responsible for all message routing, storage nodes provide disk space for file fragments and client nodes do not have a specific responsibility but just publish and retrieve files. All files are stored in the shared peer-to-peer user space, concretely the space provided by the storage nodes. In order to put a file, it is first encrypted and then encoded into multiple fragments which are uniformly distributed among the peers. Due to the fluctuation property of such technology, Wuala needed to solve the availability problem to ensure persistent storage. This has been solved by redundancy, where redundancy does not mean replication as this technique would result in too much overhead because of the low node availabilities. Instead, Reed-Solomon error correction codes, also known as erasure coding, is used. With such an approach, any m file fragments can be encoded into n and then any m out of those n are sufficient to reconstruct the whole file. This allows Wuala to distribute all file fragments in the DHT and additionally store any m fragments on their server in order to provide an additional backup. Furthermore, every client periodically checks whether all fragments are stored properly and reconstructs and uploads missing ones in case, for example, a responsible storage node has left the network forever. The lookup of content works the other way around where any m fragments in the network have to be found, also from the server if necessary, decoded into the initial n fragments and finally decrypted. So as to query a file, clients contact their closest super node to which they are connected and which routes the request to the super node associated with the storage node holding the file. This can be highly parallelised as each file fragment can be requested simultaneously. Storage nodes contact client nodes directly in order to reply. In case a storage node holds a very popular and thus frequently queried fragment, it may also redirect the request to client nodes it recently served with the fragment, so as to preserve its capacities. The routing in the Wuala network is optimized by mixing regular and random connection graphs. Super node instances establish relations to their direct neighbors, as well as to some random super nodes further away. This allows to have both a low diameter and a high clustering and thus profit from small world effects. The routing tables grow dynamically with the size of the network and the number of passing by messages, which contain routing information similar to the ones in FreeNet, and hence allow the super nodes to learn about the network. The team behind Wuala further developed various complex mechanisms that deal with incentives and fairness within the network and thus prevent free-riding as much as possible. Due to this, each clients amount of usable online storage is derived from its own provided local disk space and its online time. Those values are measured by analytic mechanisms. Furthermore, the more upload bandwidth is provided, the more more download bandwidth can be get from a client. So as to guarantee this, a distributed reputation system that is not susceptible to false reports and other forms of cheating [3] has been applied. In order to secure the uploaded files, they get symmetrically encrypted with an 256-bit AES key on the client prior to that. Decryption is only feasible with the user password, which never leaves the client machine. This encryption also includes all the meta data of the file. Sharing of files is made possible by an asymmetric RSA friendship key that is used for authentication, each of which having a length of 2048 bits. In its early stages, Wuala was controllable over a website and later on by means of a dedicated client with a separate user interface. Meanwhile, the direct file system integration is supported for the most common operating systems.

6.3.6.3 Economical Factors

Stating facts about Wuala's initial economical intentions are hard to make, however, it should be safe to assume that the team rather focused on becoming popular and gaining

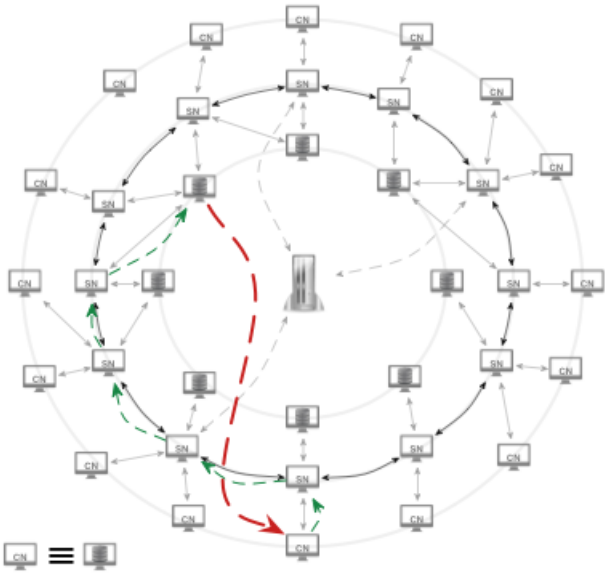

Figure 6.6: The Wuala architecture.

in numbers of users. Today, a small amount of space can be used for free and additional storage may be bought or gained through referrals. Also, LaCie tried to push the service by offering special deals, namely greater amount of storage for a limited period, when buying it hardware. It probably also rather focused on establishing relationships with business partners, as the provided privacy and encryption is an often required feature not served by common cloud storage services at the time. Evidence for that might also be the newly instantiated "Wuala Business" service, where business users can directly negotiate with the company's sales team [12].

6.3.6.4 Pros and Cons

Wuala convinces with many welcome features that were not widely spread at the time of its start. The end-to-end encryption applied to network traversing files guarantees privacy. But on top of that, Wuala also supports synchronizing, versioning and backing up of files. Its architecture guarantees a high resource and thus file availability due to the erasure codes and the fallback servers. The downside of Wuala is that it makes use of some slow mechanisms. On the one hand, the client-side encryption and decryption is very slow and might hinder the adoption of some users. On the other hand, the encoding and decoding of the file fragments consumes a lot of computer power and thus reconstruction of files takes quite some time, too. Wuala's strategy of security has also its price: once the user password is lost, it cannot be recovered. In addition, since the software is developed as closed source, nobody is able to proof that the security mechanisms actually are properly implemented. Some potential business users hence might argue that there is no evidence of the absence of a back door. Therefore, Wuala recently announced its plans to lay open the encryption-related code parts for an external audit [11].

6.3.6.5 Essential Properties

Although Wuala is implemented by means of P2P techniques and provides a solution for file sharing, it is not its main mission. The main goal rather lies on synchronization, versioning and backing up of files. The most interesting difference to other, comparable systems is the central element within a fully decentralized network, which makes Wuala

to some kind of hybrid system. The approach of having a central fallback server allows to ease the bootstrapping problem in an initial stage (comparable to Napster) and to provide "server-cloud" backups in the long-run (due to erasure codes). However, Wuala did not become as popular as other services. Maybe because the main focus is not on sharing and many similar solutions, like Dropbox or Amazon S3, are already commonly used.

6.4 Categories

The last chapter made clear that P2P file sharing services can be compared on several levels and from various perspectives. To enframe the discussion about one specific topic at the time, this report introduces categories among which separate conclusions will be drawn. Hence, not only technical aspects, like architecture and content, but also more commercial thematics, such as legal issues or monetary aspects, are covered. The following subsections specify why the study of the respective factors are relevant. Also, conclusions gained from this study's analysis are drawn and presented for each of the following categories.

6.4.1 Architectural Properties

This category pulls together the factors that indicate properties about a system's architecture. The following factors are investigated by this work and summarized in Table 6.1.

6.4.1.1 Architecture Type

The type of architecture indicates the basic building blocks of the system and primarily shows how it works. First main and essential differences can be found by comparing to architecture types. Furthermore, a systems architecture reveals the technical feasibility of many other features (e.g., searching, swarming) and also tells something about the state-of-the-art of technology (e.g., different generations of P2P).

6.4.1.2 Central Elements

The existence of a possible central element bears witness about the pureness of the P2P approach and thus might expose a possible "single-point-of-failure". By this, statements about a system's vulnerability to both technical and legal attacks become doable. On the other side, central entities indicate roles and responsibilities within the system and often support the feasibility of other features (e.g., search, user authentication).

6.4.1.3 Search

Basically, the availability of a search functionality within a file sharing system can be equated with the users needs. Often, a system is not of interest if there is no search mechanism. Also, the search of a service is an indication of how good the system provides content analysis and retrieval and provides a degree of how good requests can be handled. Furthermore, differing services implement different search approaches (e.g., index, flooding, key-based) in different manners (e.g., internally, externally, not at all).

6.4.1.4 Anonymity

Another factor by which varied file sharing applications can be compared is their support for anonymity. By this, a sign of how good the system protects its individuals, both by what they do and what they store on their disks, can be gained. Large parts of file sharing

activities are about committing illegal actions, so probably some developers try to make their system different by providing it. And even if it is not about illegal files, some users care about their privacy.

6.4.1.5 Replication

An indication of how robust the system is in terms of churn can be derived from investigating its replication technique. Replication tells about the file availability and is thus a measurement for the quality of the system in terms of persisting data loaded up to it. Furthermore, this factor describes a typical feature that emerged from the P2P technology needs.

6.4.1.6 Free-Riding Protection

By comparing different systems in terms of protection against free riders, some basic principles like dealing with fairness or load balancing can be revealed. Such a protection also tells about participation incentives or can be considered as an indication of how much normal users are cared about.

	Type	**Central Elements**	**Search (Completeness and Speed)**	**Anon.**	**Repl.**	**Free-Riding Prot.**
Napster	client-server	yes	complete, fuzzy, fast	no	yes	no
Gnutella	P2P, two-tier	no	incomplete, slow	no	yes	sometimes
KaZaA	P2P, two-tier	no	incomplete, slow	no	yes	sometimes
FreeNet	P2P	no	no search	yes	yes	no
BitTorrent	P2P	no	no search	no	yes	yes
Wuala	hybrid	yes	complete, fast	no	yes	yes

Table 6.1: Architectural Properties Category

6.4.1.7 Conclusion

The study described by this document highlighted that pure P2P-based systems are more safe from legal issues and shut down than other approaches. This does not mean that every central entity should be removed, but rather that having a central entity increases the probability to get attacked. Another conclusion describes the search, whether exhaustive or not, to be an important factor for a system's success. Users want to explore the network in order to find concrete files to download and thus the search is an essential factor. An exception to the case is FreeNet, where searching is disabled by intention. This, on the other hand, shows that searches actually might exist in many forms and depend on the user needs: built-in, external or even disabled. A further observation reveals that search approaches can be implemented to be fast, complete and fuzzy in client-server systems, but are rather slow and incomplete in P2P-based solutions. Anonymity seems not to be crucial for the success since the general perception is that individuals seldom get pursued. Furthermore, an anonymity requirement would directly contradict to search requirements as IP addresses lead directly to the providers of content. Another observation of this study concerns the free-riding protection factor and shows that people mostly do not care about cadgers in the system but are rather glad to use a hacked client (i.e., KaZaA Lite)

as long as they get their stuff. On the other hand, developers must care about it and protect their system in order to balance it and guarantee a success in the long-run user experience.

6.4.2 Content

File sharing applications also show differences in terms of the content they deal with. This category therefore concentrates on content size, diversity and quality factors. A complete comparison is shown in Table 6.2.

6.4.2.1 Content Size

The supported file size is a system property and gives enlightenment about the architecture, the network overload and the scalability of both network and storage. The content size might have a high impact on the traffic and hence on the underlying network. Also, the support of bigger content size directly influences the content diversity and increases it.

6.4.2.2 Content Diversity

The content diversity describes the broadness of supported file formats. Some systems concentrate on specific files while other allow any type of content or media, including music, films, software and games. The diversity is directly influenced by the supported content size, as the allowance of bigger files introduces the possibility of sharing bigger formats. The diversity factor might serve as an indicator for the systems modernity by respecting the state of technology and the user demands.

6.4.2.3 Content Quality

The quality of a systems network content tells how good the individual files are and whether they really are what they claim to be. The content quality also serves as measurement of how much junk, spam or malware is distributed. Overall, it tells about the user satisfaction and the availability of a possible system moderation. It further tells about the community and its support of the system (e.g., moderation, rating).

	Content Size	**Content Diversity**	**Content Quality**
Early Napster	small, mp3 files only	limited	medium, no content checks or moderation
Gnutella	unlimited	unlimited	medium, content checks but no moderation
KaZaA	unlimited	unlimited	medium, content is identified by content hash
FreeNet	unlimited, junked files	unlimited	medium, loaded by known content-key
BitTorrent	unlimited	unlimited	good, active swarms and moderation/rating
Wuala	up to 100 GB per file	unlimited	good, own and shared content

Table 6.2: Content Category

6.4.2.4 Conclusion

To conclude, all the factors, content size, diversity and quality, contribute to the overall user satisfaction by making the system more attractive. They are all important in the long-run competitive market and thus for the success of the system, since all of them reflect emerging user needs and requirements. Additionally, the content quality serves as a measurement for both the spent community effort and the general will to actively participate, provide and share qualitative content. It is thus an indirect success factor.

6.4.3 Laws

This category focuses on legal aspects and consists of only one factor: Legal issues. Because file-sharing systems are uses by individuals to exchange copyrighted material, the content industry has a high interest in preventing the use of these systems. Therefore they try to legally forbid the system and thereby forcing the operators to shut down. Under this factor we analyze if and how often a certain system was confronted with legal issues.

	Legal Issues
Early Napster	serious, shut down by RIAA
Gnutella	depends on client: LimeWire prohibited, FrostWire as successor
KaZaA	Several lawsuits right from the beginning, finally brought down
FreeNet	No legal issues known so far, but some countries may prohibit the use of Freenet
BitTorrent	The system itself (BitTorrent.org and BitTorrent.inc) never had court proceedings
Wuala	none, focus on synchronization and backup

Table 6.3: Law Category

6.4.3.1 Conclusion

The first three systems were targeted by lawsuits. Napster is special in two ways: It was the first system of its kind and not build with the possible thread of being shut down by law in mind and the central server offered the perfect target to easily shut down the whole system as well as the source to prove, that actually illegal content was shared. In the case of the Gnutella protocol the lawsuits targeted different Clients. While LimeWire was forced to shut down, FrostWire is still available. Sherman Networks which run KaZaA were targeted heavily by lawsuits and final forced to surrender. Because the KaZaA client was the only one which kept the FastTrack network up and running (no free-riding) the network collapsed with its shutdown.
In order to be a target for lawsuits a system needs to provide the following some of the following attacking points: A certain amount of users is needed. The content industry is focusing on the systems with the biggest user base first "big fish first" working its way down to the smaller ones. Another point is that a system needs to provide a mechanism to reveal its content. Each of the first three systems provided a search function, which made it possible to prove that illegal content was available in the network. Finally central elements (like servers) provide a good target and make it much more likely to be attacked.

6.4.4 Monetary Aspects

This category focuses on the monetary income streams of the different systems. If a system is able to generate money for its developers and operators it is more likely to be

maintained and extended in the future and because of that survive future challenges. A complete comparison is shown in Table 6.4

6.4.4.1 Cost (of client)

The cost describes the selling model of the client software. Often software is sold to the user over a licensing model bound to a certain version of the software. This factor focuses on the question if and how clients are sold to the users.

6.4.4.2 Monetary Revenues

This factor focuses and explains the different recurring monetary income channels i.e., donations, money through advertisement, monthly fees etc.

6.4.4.3 Business Partners

This factor focuses on connections in the form of contracts to other companies. Such contracts may ore may not lead to direct cash flows but, even more important, show if a system is standing alone or is actually connected to other companies and industries. Business partners may indicate a stronger and more durable business model, which may be plosive for the survival of the system.

6.4.4.4 Business Users

This factor answers the question if other companies use a certain system or protocol, either internally or externally. If a system has business users the chances of durability and improvement are increased, because of interested users and potential supporters.

6.4.4.5 Conclusion

Clients today are free which causes a high pressure on new clients to be free as well in order to attract users at all. The monetary revenue channels differ from system to system and distinguishable feature. While some rely purely on the community (Freenet) others use advertisement or even contracts to third parties. The content industries as business partners is quite a common model while only BitTorrent and Wuala have reported business users.

6.4.5 Quantitative Indications

In this category fall all measurable and quantifiable factors of the different systems. A complete comparison is shown in Table 6.5

6.4.5.1 Number of Competitors

This factor gives an overview of the number of other systems available at the same time, which may indicates the amount of competition and needed effort to attract new users and keep existing ones.

6.4.5.2 Number of Clients

This factor focuses on the amount of different clients for the same network and protocol. Each client is different designed for a specific user group by offering features and functionalities. Therefore the more different clients exists the more potential users may be attracted.

	Costs (of client)	**Monetary Revenues**	**Business Partners**	**Business Users**
Early Napster	free	none	none, just with reincarnation	none
Gnutella	freeware, shareware, commercial (depends on client)	only from shareware and commercial clients	none, protocol can be used free of charge	protocol can be used free of charge
KaZaA	free but bundled with adds and malware, free third party clients	built in channels for advertisement, original client bundled with malware and toolbars.	After the lawsuits they cooperated with the content produces (industry).	none
FreeNet	Free for everyone, donation possible	Only over donation	No direct partners	none
BitTorrent	free, several different client working on the BitTorrent protocol available. BitTorrent Inc. provides to clients: free and a plus version	Third party sites providing the .torrent money through advertisement. BitTorrent Inc sells their plus client as well as support to business customers	Hollywood studios, Venture Capital Partners and several different partner like ACCEL, dcm, DAG and more	Blizzard, Facebook, Twitter
Wuala	free for up to 5 GB, license model for more space	only from licenses	LaCie	many, but no specific published

Table 6.4: Monetary Aspects Category

	# Competitors	# Clients	Estimated Users @peak	System Lifetime
Early Napster	none, was the first	one	26.4 millions	1999 - 2001 (2)
Gnutella	many, diverse Gnutella clients, eDonkey2000, FreeNet, BitTorrent	many	3 millions	2000 - today (13)
KaZaA	many, Gnutella, eDonkey2000,...	One official and several hacked third party ones.	4.2 millions	2001 - 2012 (11)
FreeNet	none, only system focusing on privacy so far	Freenet client, Frost (newsgroup client build on top of Freenet)	no number available	2000 - today (13)
BitTorrent	few - only system focused on the distribution of large files	More than 20 different ones, not all available on all platforms	150 millions	2001 - today (12)
Wuala	many, Dropbox, Amazon S3, etc.	one	not available, but small	2008 - today (5)

Table 6.5: Quantitative Indications Category

6.4.5.3 Estimated number of users (@peak)

This number shows the estimated amount of users at the peak of the system. The number of users is of high importance for peer-to-peer based file-sharing systems for several reasons. Each user provides a node to the network which means the more user the more stable the network becomes. The users are also the actual (network)content providers which means the more users the higher the potential content diversity. Finally the users build the community around a network and the bigger this community gets the more likely new users will be attracted and the stronger the support for the system itself becomes.

6.4.5.4 System lifetime

Simply shows the time span in which the was available and used.

6.4.5.5 Conclusion

The number of direct competitors (offering the same functionality) seems to have an impact on the number of users of a system. Most users only use one system at the time and therefore, the total number of available users is divided among the different systems. For example Napster was the only system of its kind at the time and combined all users in its system, while Gnutella and KaZaA are direct competitors and attract only a part of the whole file-sharing community. All systems in the competitive environment (Gnutella, KaZaA and BitTorrent) offer several different clients. However it is not clear if this leads to bigger success of the system or is just a side effect of the community.

6.5 Summary

The last chapters stroke and denoted parts that seem to shed some light onto the reasons why file sharing services are able to reach apparent successes. This chapter delivers a brief

recapitulation on those parts and tries to explain them in more detail. In addition, it integrates statements about properties that indeed might be among the essential features of such sharing services.

6.5.1 Evidences of Success

When asking for the reason of success of P2P file sharing platforms, there is one very intuitive answer: people are able to get content for free. The sheer fact that content can be copied and distributed with almost no cost, at least once it is digitized, procured a big boom for such services. Also, new technical approaches to deal with digital files, such as junking and hashing, emerged over time and supported the distribution of content around the globe even more. The implication of such sharing techniques being used to distribute copyrighted content is pretty obvious when comparing the two possibilities to get it. A user either goes to the physical store and looks for the product - it might not even be in stock - and then pays for it, or she quickly installs a free P2P file sharing client, queries a search around the network and downloads the content without any charge. Hence, this work concludes that the discussion about the success of file sharing is highly controversial and most known platforms record major successes due to piracy. On the other hand, the sharing of illegal content causes big problems for all those platforms which put their system into a morally gray area since massive lawsuits have to be dealt with. Today, BitTorrent is one of the most successful sharing platforms, as the system only provides the technology to efficiently share data and leaves no room for legal attacks. The well-known .torrent-files are exchanged through third-parties, specialized in collecting, moderating, indexing and providing easy search functions for the clients. Interestingly enough, those third-parties are able to generate a lot of money - through advertisements - whilst serving as the target for legal claims. Some kind of symbioses can hence be observed. This does not mean that the P2P system is not successful itself, but simply that it is able to survive as a platform. Such a survival, however, is also bound to some other factors. By "being able to survive" the endurance describes the ability a system shows by facing competitors that may even support more sophisticated technology or features. A main prevention factor is to accumulate a large user base as fast as possible. Clients do not tend to leave a service they are already used to and rather seem to ignore the fact that another service might be better. Therefore, new services that are the first of their sort tend to acquire large successes when it comes to user adaptation. Also, a system with more participants is able to provide even more files that can be queried and exchanged. This, on the other hand, kites even more users to participate, an effect which is known as "network effect". Furthermore, the architectural design principle of P2P deals with the scalability of the system and the continuous enlargement of the network does not prevent a problem as compared to a server-based technology. However, various implementations can be faced and not all approaches deal equally well with the prevention of bottlenecks. It has been shown, that the intrinsic attributes of P2P approaches, like decentralization and scalability, are major success factors of file sharing services. Other known systems, like MegaUpload or Rapidshare, suffered from enforced shutdowns or were hit by lawsuits initiated by content industries. The reason why such attacks were successful was due to the exposition of at least one central element. This report thus defines the utmost technical success factor in the ability of the application to work in a fully distributed manner. Further factors are more complicated to determine as their nature cannot be ultimately be categorized to either be purely positive or negative. For example, the ability of the system to query and search files requires a disclosure of some form. Whilst this is excellent for customers, it eases the job of initializing lawsuits as the existence of content can be proven.

6.5.2 Desirable Properties

This study reveals many factors that influence the success of a file sharing platform. Some of them are very intuitive and obvious, others rather hidden and quite domain dependent. Such different domains show and introduce properties that emerged from specific user needs or reactions to time-dependent events. Hence, this work's intention is also to enlist some of the discussed factors and features that can safely be considered to be desirable to the greatest possible extent. Such desirable properties could serve as indications and ideas for future systems and might help to reach a certain level of adaption. In order to highlight the properties, two different perspectives have to be respected: the user's and the developer's view. As a user, an initial attraction is the availability of free-of-charge software clients that open the gate to offers that avoid any advertisements. Such offers, concretely the range of available files, should be as rich and diverse as possible, like in BitTorrent, and any file format or content restrictions should be disbanded. Also, privacy and security, like in Wualas system, or anonymity, as provided with FreeNet, might be eligible. Such anonymity should ideally not destroy any search requirements but rather coexist with search approaches that provide mechanisms to start fast, complete and even fuzzy inquiries, such as with Napster. On the other hand, coming from a developer's view, making some money serves as motivator. Furthermore, developers want to enhance their system on a technical level so as to be prepared for high liability. Adaptation to individual hardware resources, such as heterogeneity and load balancing, or the inclusion of node localities could be of high desire in order to optimize the overall system performance. Furthermore, a fully decentralized approach that ensures a stable and reliable service, also in the long-run, is quite self-evident. Another question might deal with the assumption of responsibility concerning the commitment of illegal actions, namely the infringements of copyright. Some ways propose to delegate responsibility to the users and just provide the technology, other ways suggest countermeasures like the controlling and monitoring of the system so as to protect it. Unfortunately, some desirable properties conflict at least to some degree and might be rather hard to resolve. The provision of free software, for example, and the chance to earn money from it at the same time seem to contradict. The same applies for privacy, security and, worst of all, anonymity that all make any kind of user registration impossible. And after all, how should fast, complete and fuzzy searches be possible when relying on an architecture without central entities that would normally assure a high network performance? However, the implementation of a system that provides the simultaneous existence of such conflicting properties might serve as interesting research and development areas and promise success in the future.

6.6 Conclusion

This section briefly summarizes and reflects the topic of distributed file sharing by giving an overview and a comparison of different platforms. Along this, the conclusion of this report points out the most mentionable findings. Finally, some propositions about potential future work is specified.

6.6.1 Conclusion

This report revealed that answers to the reasons of success can be found in various parts of the system. Each discussed category already unfolded diverse conclusions and the fact, that a system has to go with the trends of the time it is developed and used in is crucial for its prosperity. Overall, this report concludes that the success of file sharing applications is mainly based on the fact that users get enabled to access digital goods without charge and therefore any system's main focus needs to lie on protection. Concretely, ideally no

central elements that might act as a "single-point-of-failure" should be used. Anonymity of both providers and requester of data should preferably be guaranteed at all times. However, such a property is hardly feasible in case easy and powerful search systems need to be in place so as to query the whole magnitude of the network and eventually find the correct result in moderate time. Further, protection against pollution and junking should be supplied. Besides factors like content size, diversity or architectural properties, more oblique factors are not really influenceable by system creators. To be concrete, the community might serve as a good example: even in case a system's technical attributes are not optimal, a broad and willing user base can compensate such a fact by helping, providing, protecting, cleaning and maintaining. For new systems, however, there is another hurdle to take: users tend to stick with what they have today. During the last few years, a bunch of new systems have emerged while depredating the customers' overviews and leading to high chances that users rather stay with their current solution, even if it is not perfect, just because they are used to it. Although many systems got closed, the request after new and better file sharing applications is emerging. This raises the question whether it is time for an open system that allows for content to be shared legally and transparently. In any case, this study's factor enlistment and the compilation of desirable properties can be considered as points of reference or visions.

6.6.2 Future Work

This document shed light onto some of the most popular and therefore intrinsically most successful P2P-based file sharing services until today. Although the study of the single systems is pointing out the most important aspects regarding the respective architectures, no analysis of the in-depth details and design principles are made. Doing so would certainly reveal more low-level factors that unnoticedly influence the success of particular systems. On the other hand, more concrete factors could be found by just investigating the history of file sharing. Many historic events - such as shutdowns, court orders or changes in legal status - and developments are known to have impacted customer behavior, as well as open-source developer communities. Although this work tries to bring up and interpret some of those happenings, it is by no means exhaustive. Another promising appendage could deal with vendor-customer relationships, as proposed in [8]. In addition to that, such research could be extended to open-source projects, where no particular business instances can be identified but rather a bunch of more or less anonymous developers are involved in the creation of more sophisticated sharing applications. As privacy and trust seem to emerge more and more important in the web of today, distributed systems can as well be assumed to undergo critics of such sort. Further success factors may be found by taking the current wave of security-oriented requirements and analyze the new generation software that may emerge from it.

Bibliography

[1] Ranjita Bhagwan, Stefan Savage, Geoffrey M. Voelker: *Understanding Availability*, Department of Electrical and Computer Engineering, University of California, San Diego, 2003.

[2] Ian Clarky, Oskar Sandberg, Brandon Wiley, Theodore W. Hong: *Freenet: A Distributed Anonymous Information Storage and Retrieval System* 1999

[3] Dominik Grolimund, Luzius Meisser, Stefan Schmid, Roger Wattenhofer: *Havelaar: A Robust and Efficient Reputation System for Active Peer-to-Peer Systems*, Computer Engineering and Networks Laboratory (TIK), ETH Zurich, 2005

[4] Xinli Huang, Yin Li, Wenju Zhang, Fanyuan Ma: *Smart Search over Desirable Topologies: Towards Scalable and Efficient P2P File Sharing*, Department of Computer Science and Engineering, Shanghai, P.R. China, 2005.

[5] Jintae Lee: *An End-User Perspective on File-Sharing Systems*, Commun. ACM 46, 2 (February 2003), 49-53.

[6] Jian L,iang, Rakesh Kumar, Keith W. Ross: *The KaZaA Overlay: A Measurement Study* Polytechnic University, Brooklyn, USA, September 2004

[7] Carmen Guerrero López: *Measuring Bittorrent Ecosystems*, Universidad Carlos III de Madrid, Spain, 2012

[8] Heng Xu, Hao Wang, Hock-Hai Teo: *Predicting the Usage of P2P Sharing Software: The Role of Trust and Perceived Risk*, Department of Information Systems, University of Singapore, January 2005.

[9] P2P Music File Sharing Dropped After Limewire Shutdown, NPD Says; `http://www.pcmag.com/article2/0,2817,2382494,00.asp`, December, 2013.

[10] Requiem for Napster; `http://www.techhive.com/article/100004/article.html`, October, 2013.

[11] Wuala And the Patriot Act; `https://support.wuala.com/2013/08/wuala-and-the-patriot-act/`, December, 2013.

[12] Wuala Business; `http://www.wuala.com/en/business`, December, 2013.