University of Zurich
Department of Informatics

Burkhard Stiller
Fabio Hecht
Maurizio Lo Bosco
Peter Racz
Guilherme Machado
Andrei Vancea

Internet economics V

TECHNICAL REPORT – No. 2010.IFI-2010.0007

June 2010

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

B. Stiller
F. Hecht
M. Lo Bosco
P. Racz
G. Machado
A. Vancea:
Technical Report No. 2010.IFI-2010.0007, June 2010
Communication Systems
Department of Informatics (IFI)
University of Zurich
Binzmühlestrasse 14, CH-8050 Zurich, Switzerland

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 spring term FS 2010 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 V" discusses a number of selected topics in the area of Internet Economics. The first talk "IPTV and its Quality Assurance" discusses different IPTV technologies as well as the quality assurance and economic aspects of an IPTV service. Talk two on "Economic Challenges of Standardization" presents the process of specifying Internet standards and shows case studies on the browser war and web standards. Talk three on "Economic Aspect of Open Source Software" discusses different open source licenses, advantages and disadvantages of open source software, as well as business models in the open source field. The fourth talk "Standards and Models of Expressing and Monitoring SLAs" focuses on Service Level Agreements and different models for expressing and monitoring a service. Talk five on "Opportunities and Threats of New Technologies to P2P Networks" discusses the impact of new technologies, like IPv6 and fiber to the home, on P2P networks. Talk six on "Economics of the Digital Divide" discusses the fact that some people have effective access to digital information and technology whereas others have only very limited or no access at all, and presents different approaches for bridging this gap. Talk seven on "Economic

Challenges and New Opportunities using Clouds" provides an overview on Cloud Computing and discusses economic aspects of Cloud Computing in three different use cases. Finally, "Optimization Effects of Smart Metering" as talk eight gives an overview on smart meters that are used to record the power consumption of a household with the goal of reducing the overall power consumption.

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 Fabio Hecht, Maurizio Lo Bosco, Peter Racz, Guilherme Sperb Machado, Andrei Vancea, and Burkhard Stiller. In particular, many thanks are addressed to Peter Racz 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, June 2010

Contents

Chapter 1

IPTV and its Quality Assurance

Bo Chen, Thomas Hunziker, Evgeny Kislitsyn

IPTV is getting more and more popular. It is a new technology that has managed to enter an already well established market with very high entrance costs. This makes it especially interesting to all market participants and so it is important to understand how this technology works and what assets and drawbacks it will bring. In this seminar work we will show the differences of the different distribution technologies of TV, discuss the architectural aspects, the quality assurance of the IPTV service and the economic aspects.

Contents

1.1 Introduction

1.1.1 Motivation

In the present world we have many ways to receive TV programs. There is cable, satellite, Digital Video Broadcasting Terrestrial (DVB-T) and recently also the Internet. The most popular way is still by far cable followed directly by satellite. But there are still a lot of new technologies emerging that suggest new ways of broadcasting TV as the old ways of transmission do not satisfy customer demands in the age of Web 2.0 any more. New customer demands also generate opportunities for new technologies to enter already well-established markets, for example in the recent years the Internet Protocol Television (IPTV) technology has entered the TV market and is rapidly growing.

1.1.2 Outline

The goal of this seminar work is to give an overview of the technology of IPTV; especially its architecture, the quality assurance of it and its economic aspects. Section 1.1 provides an overview of the current TV broadcasting technologies and illustrate how IPTV will fulfill new customer demands. Section 1.2 focuses on the architecture of IPTV and discusses some architectural problems. In Section 1.3 different aspects of the quality assurance of IPTV will be explained. Section 1.4 then discusses the economic aspects of IPTV. Finally, in Section 1.5 we will summarize this seminar thesis, wrap up the main points and provide an outlook.

1.1.3 Definition of IPTV

Because IPTV is the abbreviation for Internet Protocol Television, it leads to the conclusion that IPTV stands for Television over the Internet, but is only partly true for the context of our seminar work. We define IPTV as a service that is offered by a certain company (mostly a telecommunication companies e.g., an Internet Service Provider (ISP), because they have the customers who already have the most crucial requirement, a high speed internet connection) to customers as an alternative way to watch TV programs on the TV. This technology typically requires a high speed Internet access with a certain bandwidth and a set-top-box that will operate independently. The Internet connection is also mostly provided by the same company offering the IPTV solution. In fact most ISPs provide a so called "Triple Play" subscription where a customer gets telephone, Internet and IPTV. See also Section 1.4. The set-top-box receives the data needed from the Internet, decodes it and send the processed signals to the viewing media used. Customers will have to pay a fee to the company providing this service and this service also includes some special services and applications, which we will discuss throughout this seminar work. The company providing this kind of service is also in charge of the quality assurance of the service itself, which we will discuss in Section 1.3.

1.1.4 The Four Main Distribution Channels of TV

Nowadays there are four main distribution channels for TV. These are: cable, satellite, DVB-T and IPTV. In the following sections we will give an overview of the four technologies and highlight their strengths and weakness.

1.1.4.1 Cable

Cable TV is almost as old as the TV itself and its history reaches back to the beginning of the 20th century [27, p. 23-25]. It is by far the most common way to receive TV at home. About 90% of the households in Switzerland use cable [24]. This is mainly because most households already have this kind of infrastructure set up, because by the time when TV over satellite was not invented yet, this technology was by far the most stable and best solution. In order to watch TV over cable, you just have to connect the cable from your TV to the cable box in the wall. These key benefits are the biggest strength of cable TV: Its simplicity of use and its wide spread distribution range. Another advantage is that cable TV is a very robust technology and very rarely suffers from failures. But on the other hand, if we compare the costs of cable TV and the other technologies, it is clearly the most expensive one (see Section 1.4.2.1). Other disadvantages are its rather limited variety of programs and the nowadays rather poor picture quality, although in some countries you can get high definition television (HDTV) over cable for a small premium.

1.1.4.2 Satellite

In 1980, TV over satellite was a revolutionary approach for the reception of TV [11, p. 201-203]. This technology made it possible for all people to have access to high quality TV, because all they needed was a satellite dish and a receiver for the TV. They did not need to install a cable system in their house, nor did they have to pay monthly fees. Some regions are just too sparsely populated, so it is economically not profitable to build the infrastructure just for a few households. The strength of satellite TV lies in it's availability and its big variety of TV programs e.g., one can receive Japanese TV channels in the US. Another advantage is its good picture quality. Its main weakness is the rather complicated and sometimes also expensive setup of the whole system (one has to orient the satellite dish towards the right satellites). In addition, satellite TV can fail when the weather is really bad e.g., heavy rain or heavy snow.

1.1.4.3 DVB-T

Analog TV over television antenna reaches back to 1930 [27, p. 19-20]. But since the beginning of the 21st century DVB-T has been constantly replacing the old analog technology. In order to receive DVB-T one has to either have a modern TV where DVB-T is already implemented or buy a DVB-T receiver. Though this solution is very mobile and very cheap, the picture quality is highly dependent on the weather and the variety of programs leaves a lot to be desired.

1.1.4.4 IPTV

IPTV is a very young technology for receiving TV over the Internet. It has a decent picture quality and also comes with some of the benefits that only satellite TV was used to offer. Its biggest advantages are its ability to view programs in HDTV quality and its price. Aditionally, it is also more robust against the weather changes. Though its biggest disadvantage is that one has to have a high speed Internet connection in order to be able to receive IPTV, some ISPs have already started to offer IPTV only solutions [8].

1.1.4.5 Overview of the Differences between the Four Main Technologies

Table 1.1 shows a direct comparison of the differences between the four main distribution technologies cable, satellite, DVB-T and IPTV.

Table 1.1: Overview of 4 main technologies

	Cable	**Satellite**	**DVB-T**	**IPTV**
weather proof [a]	high	normal	low	high
mobility	low	low	high	low
price	high	low [b]	low	normal
variety of programs	normal	high	low	normal/high [c]
picture quality	normal	low-high [d]	low-high [e]	normal/high [f]
installation effort	low	high	low	normal
interactive applications	yes [g]	no	no	yes

[a]by weather proof resilience against wind, rain or snow is meant.
[b]high initial investment, but no running costs.
[c]this depends on the additional packages a customer subscribes to.
[d]this depends highly on the weather
[e]this depends highly on the weather
[f]this depends on the Internet speed of a customer
[g]very new technology, additional costs involved, based on IPTV

1.1.5 Why IPTV?

Why do we still need another way of receiving TV programs if we already have 3 different ways? There are several points to consider:

- Its picture quality is better than the quality of cable TV
- It is cheaper than cable TV (see Section 1.4.2.1)
- It has a greater variety of programs than cable TV and DVB-T
- Is is more robust to weather changes than satellite TV and DVB-T
- It offers interactive applications such as Video on Demand (VoD) or an Electronic Program Guide (EPG) (see Section 1.4.1)

- It builds up on existent infrastructure

These points are clearly speaking for IPTV as a new way of distributing TV programs. Thoug, the biggest problem of IPTV is that the customer needs a high speed Internet connection, which is not always given due to either personal or technical reasons.

1.2 Architecture

In this section we will explain the general architectural setup of an IPTV system and especially discuss the different architechtural styles of unicast and multicast, and centralized and distributed servers.

1.2.1 General Architecture

Figure 1.1 shows a typical system architecture for the IPTV technology [28, p. 49].

The Super Head End (SHE) is the starting point and the main source for the whole IPTV system. It is directly responsible for:

- Gathering the content from the different programming suppliers such as NBC, CNN, Fox, etc.
- Converting the input streams into the appropriate delivery forms
- Transmitting the content to the Video Serving Offices (VSO)
- May also be responsible for preparing the content for VoD that will be stored on the video servers

Often an SHE is preparing the content for hundreds of thousands of viewers, thus, for these "very large" systems often a second SHE is built for emergency failures of the SHE (such as a complete breakdown of the SHE or failure of main functionalities), in order to prevent that hundreds of thousands of customers are completely cut from TV. The SHE is a very important part and thus needs staff to monitor its performance 24/7. After the content is sent, it will reach the VSO, which is responsible for a geographic region e.g., a city. One SHE delivers content to several VSOs. A VSO is responsible for:

- Video processing e.g., converting to another codec or bitrate
- Adding local data e.g., local tv programs or local advertisement
- Storing data e.g., tv series or films for VoD services
- Stream creation e.g., the data streams to the Local End Offices - one stream per viewer or one stream per broadcast channel (see Section 1.2.2 for more details)

After that the specific stream reaches the local end office, where the IP video services, the internet data streams and the telephone signals all come together and will be sent to the customer [28, p. 52-53]. The local end office is responsible for:

- Forwarding the streams to the right customer.
- Combining different services such as Internet traffic data, Voice over IP (VoIP) or telephone signals.

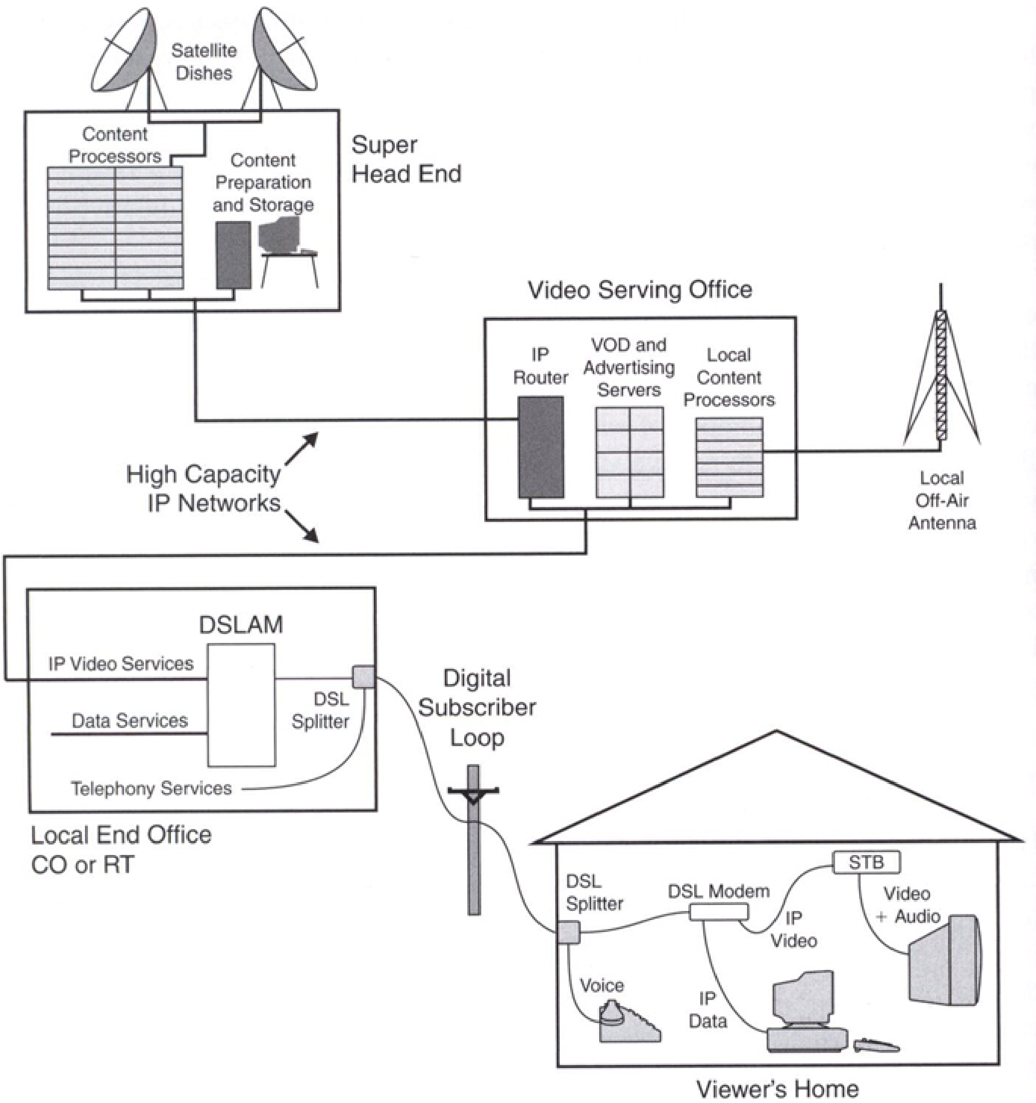

Figure 1.1: A general setup for IPTV technology [28, p. 50]

After that, the data will be sent directly to the customer. There the DSL Splitter needs to split the telephone signals from the data stream. The DSL modem will then split the different services and forward them to the right end systems [28, p. 53-54]. This is just a general system architecture scheme e.g., how it is typically or can be set up. Other possibilities are for example a partial or complete peer to peer architecture.

1.2.2 Unicast vs. Multicast

Unicast and multicast are two different technologies in order to transfer the different TV channels to the customers. In the following we will explain the 2 technologies and then summarize the main differences and the benefits.

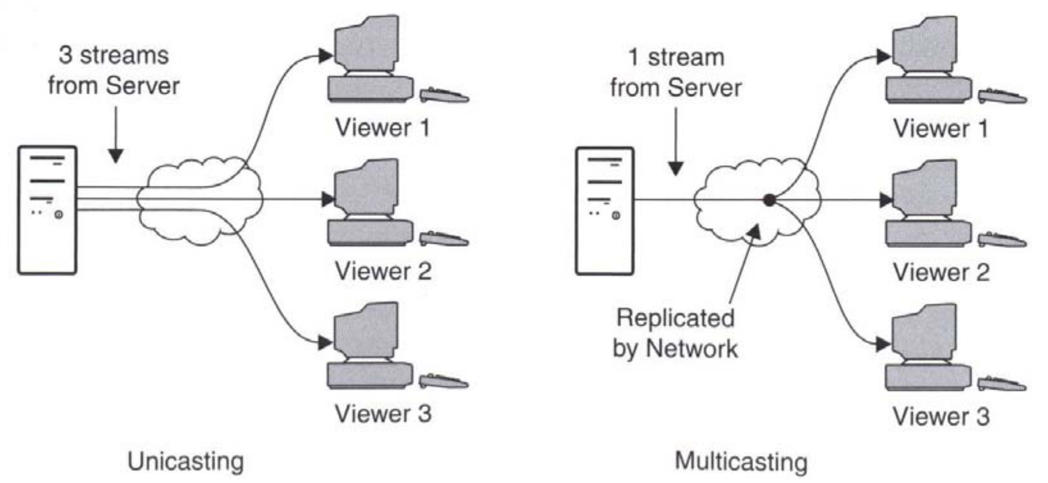

Figure 1.2: Unicast and Multicast technology [28, p. 78]

1.2.2.1 Unicast

As you can see in Figure 1.2, in IP unicasting, "each video stream is sent to exactly one recipient" [28, p. 77]. Each user has to request the video stream and then the video server has to send the specific video stream to the destination IP address. This means that if different users want the same video, for each user a different video streams have to be sent out. If there are for example 20 viewers with each 5 Mbit/s then the total capacity of the network has to be 100 Mbit/s. This may sound bad, but the most important benefit of unicasting is that each video stream is custom tailored. For some applications (like VoD) this is a most desirable feature.

1.2.2.2 Multicast

Multicasting exact opposite of unicasting. As you can see in Figure 1.2, instead of sending one stream to each user, the video stream is transferred as one stream and each user who wants to watch this stream can subscribe to it. So instead of sending a new stream from the video server to the end user for every single end user, the stream is sent to a special adresses that is reserved for multicasting [18]. The packets are then copied inside the network by the routers and sent to the receivers that have registered through a group management protocol. Mostly the Internet Group Management Protocol (IGMP) is used [28, p. 78-79]. This reduces the required processing capacity and also reduces the bandwidth needed significantly, but the users will not be able (or it will be very difficult) to send custom requests to the server, such as fast forward, rewind or pause. Another disadvantage is the rather complicated group management.

1.2.2.3 Comparison

Table 1.2 shows, that multicast is very good where a single video stream has to be streamed to wide customer base e.g., a TV channel like NBC or CNN. In contrast, unicast is good where the user has to be able to control the video flow or where only a very small amount of users watch the video simultaneous e.g., VoD services.

Table 1.2: Comparison between Unicast and Multicast

Unicast	**Multicast**
one stream per user	one stream per broadcasting channel
custom requests possible	custom request not possible
bandwitdh consumption increasing very fast	bandwidth consumption increasing very slow
processing power from video server needed	group management needed

1.2.3 Centralized vs. Distributed Server

The centralized and distributed server architecture are mainly interesting for services like VoD, where a trade-off has to be made between complexity of the system and the performance of it.

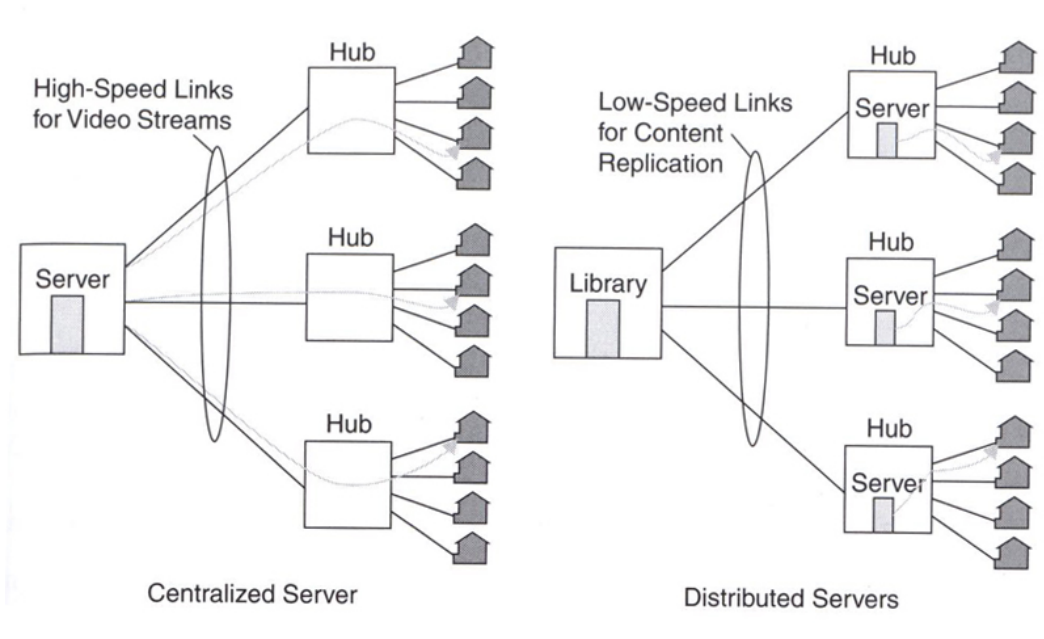

Figure 1.3: Centralized Server and Distributed Server Architecture [28, p. 115]

1.2.3.1 Centralized Server

Figure 1.3 shows that in the centralized server setup, there is only one server for several hubs. One hub is responsible for several users. If a user wants to start a service like VoD, the hub will forward his request to the central server and then the central server will send the video stream to the hub, which will forward it to the user. This obviously requires strong processing power of the server and high speed links from the server to the hubs.

1.2.3.2 Distributed Server

Figure 1.3 shows that in the distributed server setup instead of the central server, there is a library and along with the hubs there are servers. If a user requests a service like VoD in this particular setup, then the request will be directly processed by the servers at the hub. If they happen not to hold the requested material because it is rarely requested, then they will request the material from the library.

1.2.3.3 Comparison

The centralized server setup requires one powerful server that can handle thousands of requests and a high speed link between the server and the hub that can handle the traffic of thousands of requests at the same time. The distributed setup on the other hand requires several smaller servers, a library and only a low speed link between the library and the servers. This setup makes the management more complicated than just managing one server. In order to decide which setup to use one has to analyze what capabilities are given, what are the viewing habits of the users and what system architecture has been deployed.

1.3 Quality Assurance Aspects of IPTV

In this section we will discuss the aspects of quality assurance in IPTV and give a definition of quality assurance, quality of service (QoS) and quality of experience (QoE). The reasons wherefore quality assurance is important for the service provider and for the customers. To explain how we can measure the quality of IPTV, a framework for measuring QoS and QoE will be given.

1.3.1 What is Quality Assurance?

The target of this section is to give a deeper understanding of quality assurance. The main thing what a service provider must do is to give his customer a good experience in watching IPTV. The measurement of the customer's experience is not so easy. Therefore the service provider is looking for a good service quality, the general opinion is that this leads to a better quality of experience. [16].

1.3.1.1 Quality of Service

“QoS is a collective of service performances that determine the degree of satisfaction of a user of a service. QoS parameters are user oriented and are described in network independent terms.“[12]

In IPTV QoS is the ability to guarantee a certain level of video quality and sound quality. These parameters depend on the underlying network. Common network problems for IPTV are jitter, latency, bandwidth and packet loss.[2]

The bandwidth problem can be solved by using a broadcast service. A broadcast or multicast needs only one stream for all users as explained in Section 1.2.2.2. If a VoD service is desired, then a unicast is essential at least to a certain degree.

The latency problem can be avoided by using a buffer. The latency will be not smaller, but the impacts on the QoS will be smaller. For example the customer what to see the last minute again, because a phone call comes in. Then he can press the playback button, and the video stream comes out of the buffer instead loading it again from the server. So a buffer can reduce the impact of latency to the QoS.

This can be used for broadcast streams and video on demand streams. But by video on demand, the latency should not be too big, or the user will recognize this. The main problem here is to determine the needed buffer size and needed hardware memory. This is also a question to the cost for the receiver.

The jitter problem has certain similarities with the latency problem, the arrival times of individual packets is different. This can cause problems, because packages are missing at runtime. This problem can be solved with a jitter buffer at the expense of additional run-time and additional hardware.

The packet loss problem can be solved by retransmitting the packets or by using special coding algorthims. If with the help of algorithms, the problem is not resolved, the packets must be sent again. This may consume additional bandwidth.

In order to determine, for example, the buffer size a kind of intelligence is necessary, which controls the key parameters. Such a solution (framework) is explained in Section 1.3.4.

QoS and QoE try both to measure the quality from the customer’s view of point and if possible to improve it.

1.3.1.2 Quality of Expirience

“Quality of experience (QoE) describes the overall system performance from an end user’s perspective and is related to service layer and the ability of a service to meet the end user’s expectations.”[21]

For the customer the video quality and the sound quality is the important thing. He does not care about network problems such as jitter or packet loss.

The measurement of QoE is not as simple as the measurement of QoS, because the experience of the customer is not only determined by the transmitted quality of video (determine primarily by the network performance).

For example, a disturbing noise from the speaker of the TV, can strongly influence the QoE. This error can only be detected when a microphone is placed in the room, where the TV is placed. In a perfect world the service provider would install a loopback from the microphone to the service office for examine the QoE. This would consume much of the limited bandwidth, there is a need to find another way to measure the QoE. A possible method for measuring the QoE is described in Section 1.3.3.3.

1.3.2 Reasons for Quality Assurance

An interesting aspect of quality assurance is not described yet: Why is the quality assurance needed? Primarily there are two stakeholders, first the customer and second the service provider. In the next two sections the reasons from the point of view of each stakeholder are explained.

1.3.2.1 Customer

In the past there was a rapid development of the end devices especially in video quality (HDTV) and sound quality. To use these new features the quality of the source streams have to be coherent. This means that not only the quality of the video itself is important, it is also important that the delivery of the stream does not affect the quality (see Section 1.3.1.1. Therefore the customers need quality assurance.

1.3.2.2 Service Provider

From the customers point of view there is no reason to switch to IPTV for the normal TV program, execpt for VoD. But real VoD services can only be provided by IPTV because all other technologies are broadcast services.

If there is no reason for the customer to switch to an IPTV technology the service provider must give an additional benefit to the customer (for example VoD or lower price). Additionally he must assure the same quality as conventional TVs to the customer or the new technology will not be sold.

These reasons lead the service provider to invest time and money on implementing quality assurance into IPTV products. In the next section such a framework for such an implementation will be presented and described in detail.

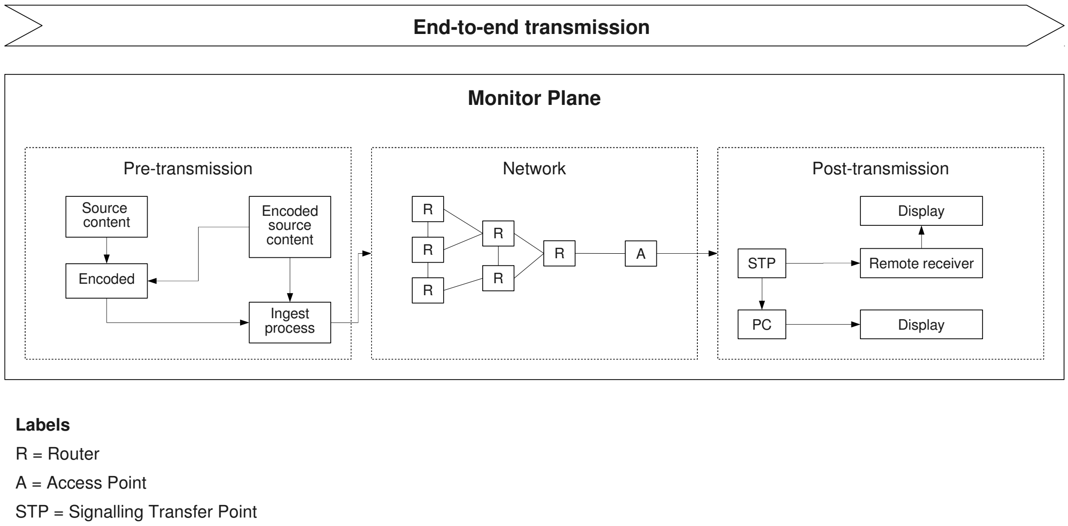

Figure 1.4: Extended IPTV QoS assessment framework proposed by *Standardization Activities in the ITU for a QoE Assessment of IPTV* [2] [1]

1.3.3 Measurement Framework for QoS and QoE

The measurement of IPTV quality is the main challenge of quality assurance in IPTV. A simplification can be done by using a framework. In this section we discuss a framework proposed by A. Takahashi et al. [1] and we extend this framework with the enhancements proposed by B. De Vleeschauwer et al. [2]. Figure 1.4 shows the framework proposed by A. Takahashi et al. It shows three zones in the transmission which should be measured. The pre-transmission layer is on the service provider side, the "network" layer is the network between the service server of the provider and the customer and the post-transmission layer is the playback of the media on the devices such as TV or PC.

1.3.3.1 Pre-Transmission

Pre-transmission is the phase before the video stream is distributed over the network. In this phase the media can be checked and some error concealment methods can be applied to the media [1].

Usually, to transmit the media over the network, the application encodes the content. For checking the media quality after the encoding there are some pixel-based methods: Full-reference, reduced reference and no-reference methods [1].

In the full-reference method the source media is compared to the encoded media. These methods are usually computationally intensive. The no-reference method is a blind test of the encoded media because there is no need for the source media (reference media). This method is usually not so computationally intensive, in some case the quality assurance is not so high. In between the reduced reference method is a mix of the other two methods and the benefits of both are used. For this method only a part of the source media is needed [1].

In practice, the no-reference method is preferred, because the quality assurance must be performed in real time. Furthermore, sometimes the source is not available. In such a case full-reference or reduced-reference method is not useful [1].

1.3.3.2 Network

There are many measurement methods for network performance. They can be categorized in passive and active methods and they can be performed in live or in test networks. Passive measurment methods are methods that use only packets, which were tansmitted anyway. Active methods are methods that send additionaly packets for analysing the network performance.

The problems described in Section 1.3.1.1 typically do not occur in the test environment, so only live systems can be used. For example, to detect a packet loss, a significant amount of packets must be send over the network. Normaly the bandwidth is already exhausted and in this situation sending additional packets is not sophisticated. Therefore only the passive method is relevant in IPTV systems. The conclusion is that for network measuring in IPTV normally a passive method in a live system is used.

1.3.3.3 Post-Transmission

The network performance (the problems mentioned in Section 1.3.3.1) and the delivered and encoded media is enough to determine the QoE. The QoE is defined as the experience of the customer. This monitoring does not guarantee the quality on the screen of the TV; consider problems with the receiver or with the cable between the network access and the receiver. These problems appears after transmitting the media (Post-Transmission). There is a need to investigate deeper on this point, to find a way to measure also the shown quality. Probably the best way to determine the QoE is to measure quality of the displayed video.

As earlier mentioned, there are generally three (full-reference, reduced-reference and no-reference methods) sorts of algorithms to determine the quality of a video stream. The full-reference method is not an option to determine the quality on post-transmission side, because there is not enough bandwidth [1].

For the reduced-reference algorithms there is at least an additional channel needed to transmit some reference parameters. For both, full-reference and reduced algorithms, the source and the encoded video stream must be closely aligned. A no-reference method does not need any additional channel and bandwidth, but there is no ability to link the source quality with the target quality. If the source quality is not uniform, the output of the quality measurement is useless [2].

The ability to measure the QoS and QoE does not have any impact on the quality itself. For example to change a parameter to increase the bandwidth, there must be some sort of application to control this. In the next section an extension to the described framework is given. This extended framework proposes a solution to this leak.

1.3.4 Industrial Application of Quality Assurance in IPTV

With the described framework the measurement of the quality is possible, but this has no positive effect on the quality itself. There is a need to apply this information to the application to improve the quality. The paradigm of knowledge plane is known as an automatic detection and recovery tool in extended networks [2]. These paradigm applied to the IPTV scenario is shown in Figure 1.5. There are two layers: Monitor plane and knowledge plane. The monitor plane is described in Section 1.3.3. The focus on this section lies on the knowledge plane and the interaction between these two layers.

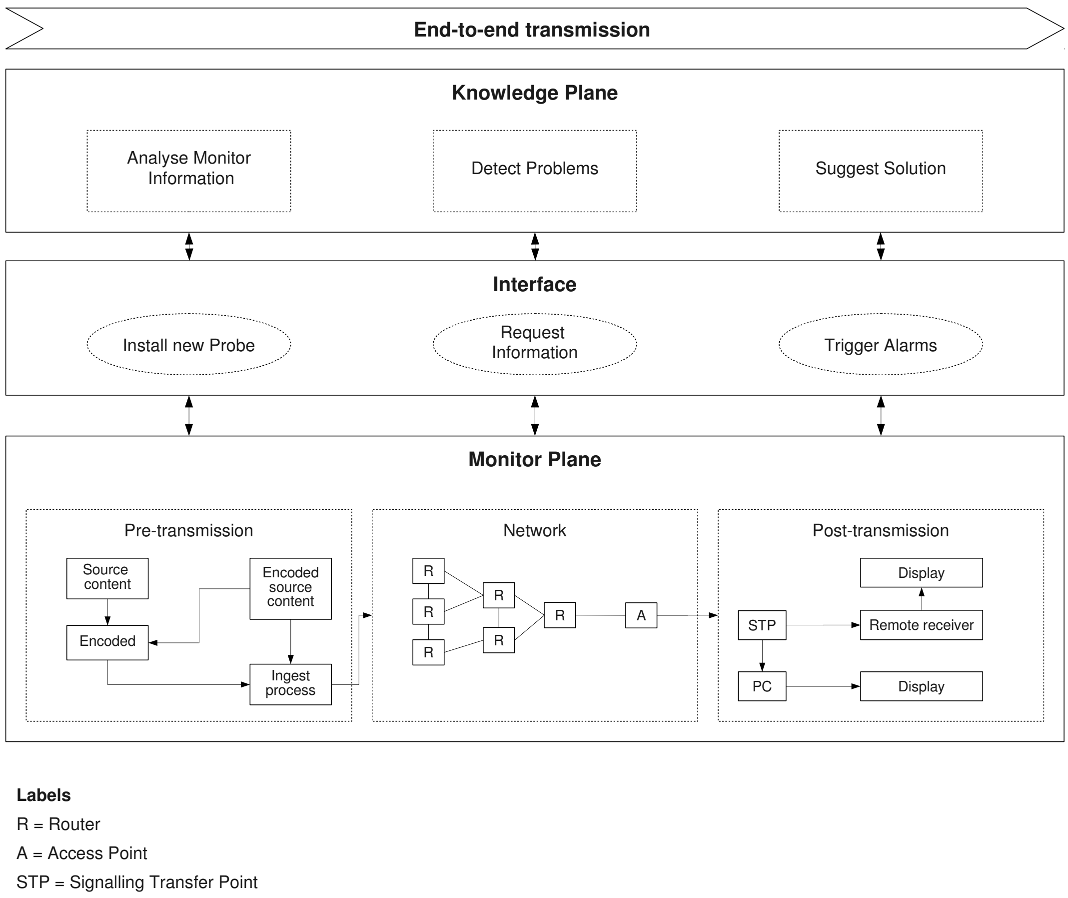

Figure 1.5: Proposed framework including knowlege plane. [2] [1]

1.3.4.1 Knowledge Plane

The knowledge plane lies on the top of the monitor plane. Its primary function is to interpret the data collected by the monitor plane, to detect some leaks of the quality and the cause of these leaks. The knowledge plane should not only receive the information from the monitor plane, it should also install new probes and new observers. We will cover the details in the following use case.

1.3.4.2 Interaction

This use case explains how the interaction between the monitor plane and the knowledge plane can be done and how it can be used to increase the quality.

The customer has installed his new television equipment with all receiver hardware. Then he starts the television and the hardware delivered with the framework described in the section above. Everything seems to work, but when the customer changes the channel, it consumes too much time. The monitor plane detects this problem and alerts the knowledge plane. The knowledge plane analyzes the problem in conjunction with additional information from the monitor plane. A possible cause can be for example a miss configuration of the firewall or the buffer size is too small. This information can be send to the service provider. If it is possible there is an application, which can correct the error or a technical person solves the problem by hand.

1.4 Economic Aspects of IPTV

Figure 1.6 shows that IPTV is a big market with over 45 million subscribers in 2010. The revenues worldwide back in 2006 were about 700 millions, the expected revenues for 2011 will be about 23 Billions [15]. These ratios shows the precipitous growth of revenues worldwide and the trend is continuousness. But like in any other new technology IPTV has a few aspects to fulfill in order to become not only a favorable but also a profitable resource [6].

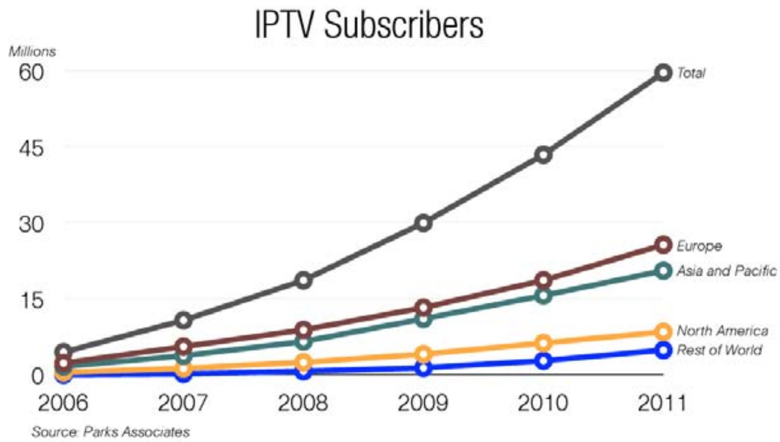

Figure 1.6: IPTV Subscribers [3]

In the following section we will go through some facts about IPTV on the world markets and its economic value. Furthermore we will describe some economic architectures, which have different ascendancy on different IPTV markets. In the past few years IPTV has shown a great growth rate within many countries. Since 2002 when Sasktel were the first commercial IPTV provider in US, IPTV became more and more interesting for the industry. It is not only the enormous growth of technology in internet bandwidth [4] through the past few years, which pushed IPTV to be the next generation television. The fiber industry, which experienced a huge boom in the past few years, was also a factor that indirectly pushed that trend. The stagnating market of cable TV and the less available technology potential gave the industry the chance to redesign the experience of watching TV. The two main drivers of IPTV are tripple play and the extended broadband technology, which allows to stream movies and TV-shows in high quality like DVD or even High Definition (HD).

1.4.1 Key Benefits of IPTV

Besides the revenues and the profit, IPTV can offer more than just another lucrative selling product. It has considerable advantages for both sides, the one of the industry and the users side.

IPTV offers the so-called Video on Demand (VoD), which can provide any kind of video data at a certain time for a small fee. With this service the subscriber is free to choose when and what he wants to see, whether it's an episode, a TV-show he missed this afternoon or a movie he would like to see again. On the other side, the provider has the full data control of what his subscriber have watched or bought which gives him a great financial overview of its sales and a better cost control. The provider can use this data to create personalized products for his customer and offers them as special subscriptions. Furthermore IPTV technology allows it to measure real audience rates. The commercial can be personalized, so the housewife does not have to watch beer commercials or IT-service publicities.

Another main benefit of IPTV is the triple play technology that transmits three kinds of information. Data for internet, voice for telephone and video for television. But this triple play technology can be used in many different ways. The data part could be used for interactive television and real-time voting in a TV-show. It could be also used for applications, which run on the media receiver. These applications could be small programs, which would extend the possibilities of television. Alcatel is offering an application, which allows the user to create an own TV-channel. This channel would be like a virtual living room where people can meet together and watch television. Furthermore the user can share his personal content like photos, videos or documents in this digital channel, and all people who are invited can see this content. The possibilities of these applications are nearly boundless. Even though the industry is in the origin of this technology, many experts believe in a major breakthrough in a few years.

1.4.2 Switzerland

Switzerland as a country of wealth and prosperity is one of the most desirable market for IPTV. The well engineered network infrastructure and large amount of potential IPTV users makes this market very attractive to roll out this new technology and to bring it to the people. That is why IPTV has shown a great growth in Switzerland through the past few years. Swisscom, that is the biggest internet and phone provider in Switzerland and started IPTV back in 2006 [25]. At that Time the IPTV technology were not fully developed.

Swisscom had a lot of problems with the roll out of IPTV [17]. On the one side was the great technology with a huge potential on entertainment technologies and on the other side there was the tremendous cost in implementing this System and running it stable. The costs per user of 1400 CHF has shown that there was a difficulty of making this technology profitable. But time has shown that IPTV is profitable: In 2010 Swisscom counts 275000 user of IPTV and the expectations are up to 400000 for the end of the year [26].

1.4.2.1 Cost Comparison of IPTV and Cable in Switzerland

Table 1.3 shows how the costs of IPTV and normal cable connection vary. But keeping an eye on the prices for TV over IPTV we can see an opportunity in every situation. Furthermore IPTV is even better, because it offers more TV programs in its basic offer and the record function do not cause additional cost for the subscriber like the cable offers do. But note that the costs for the telephone connection in the IPTV package are excluded for this comparision as they are sunk costs if we assume that a normal household needs a telephone connection anyway.

Table 1.3: Cost Comparison [5]

	Cable Costs	IPTV Costs	Savings
Normal TV (40 Channels)	26,45 CHF	21,25 CHF	(+) 5,2 CHF
Extended TV (100 Channels)	32,45 CHF	21,25 CHF	(+) 11,2 CHF
Extended HD TV	42,45 CHF	30,75 CHF	(+) 11,7 CHF
Extended HD TV, Recordable	52,45 CHF	30,75 CHF	(+) 21,7 CHF

1.4.3 Germany

With 40 million households (these days), Germany has a huge market potential for rolling out the IPTV technology [9]. In fact T-Online made the first step in 2007, by offering the first media receiver to watch TV over internet without having problems to use the personal computer online. Alice TV followed the same year, realizing the big potentials of this technology [10]. Establishing this new technology T-Online created a new market in Germany. It has in fact the same qualities like the cable TV, satellite TV, but IPTV has a bigger potential in creating a new way of entertainment, like interactive television.

Analyzing the revenues of the industry in Germany in Figure 1.7, shows that IPTV is not skimming the revenues of other companies. Through the new opportunities of IPTV,

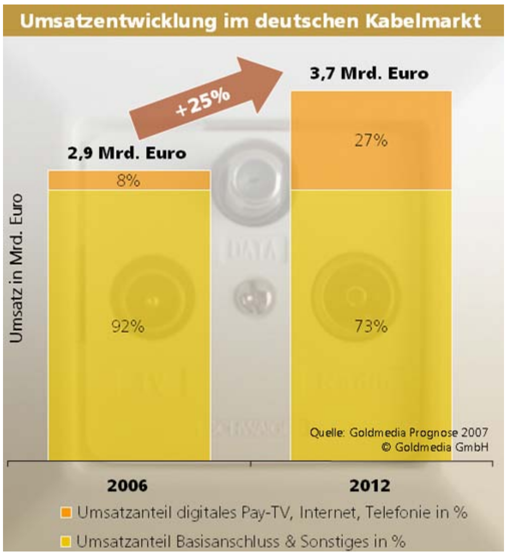

Figure 1.7: IPTV revenue forecast [7]

it generates additional revenues and takes an important position in Germanys market by earning more importance.

Nowadays there are about 15 million users of IPTV in Germany [13]. Experts are expecting a gain of 80 percent up to 27 million till 2013 [13]. Many companies realized this potential of IPTV, for this reason experts are estimating a large growth in IPTV providers for the next few years. Although IPTV were not generating that much profit in the past, the German industry believed this technology will be the next generation TV technology. The expectations of revenues for 2011 about 39 billion EUR, confirms their believes.

1.4.4 China

The GDP of China has shown in the last few years an enormous accretion, equally it reflects in the growth of IPTV services. Figure 1.8 shows that China is the biggest and the one of the fastest growing IPTV markets with over 172 million of cable users [22], and about four million IPTV users.

The government pushed the industry by giving the population subsidies of up to 13 percent per new household appliances. In addition to the help of the government the industry is spending more and more money for new investments in IPTV innovation and technology. In 2006 the investments in IPTV technology were about 6.3 million USD. In 2008 the investments were multiplied by nearly four and raised up to 22.7 million USD [19].

1.4.5 Digital Rights Management

In the past few years the movie industry had many problems in selling their products through digital rights. Widespread piracy all over the internet caused an enormous loss in revenues in the movie industry. The Motion Picture Association of America estimates

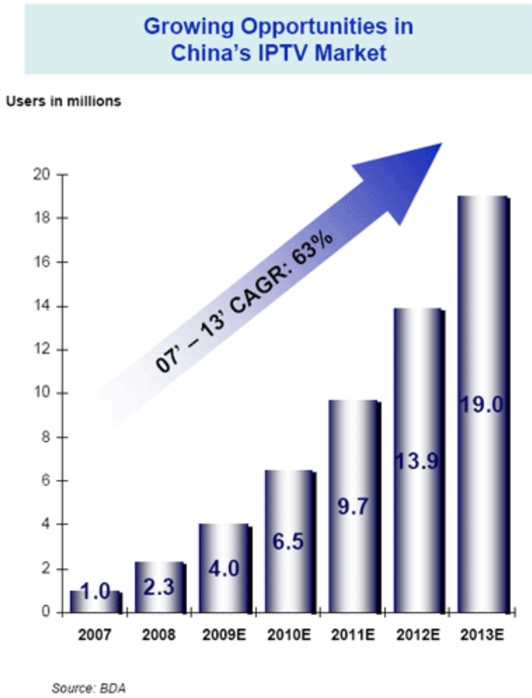

Figure 1.8: IPTV growth in China [22]

a loss of revenues of about 3.5 billon USD [14]. The DVD itself cannot be protected from being copied and distributed all over the worldwide web. This problem could be solved with IPTV technology. Unlike the DVD, which have no protection at all, watching videos through IPTV is through a security locked internet connection. Hacking or any kind of trying to get the signal of this transmission is quite impossible without knowing the right key, making this connection secure and avoiding possible fee loss through piracy copies [29]. Experts are speculating about the relief of DVDs through VoD.

1.4.6 Suggestive Economic Networks

The technology of bandwidth has been changed in the last ten years and it is still getting faster. But the consistent growth of the number of IPTV subscriber needs more than just a fast internet connection. Broadcasting TV is easy to handle and not that complex as delivering VoD as a unicast. The exact problem is not the connection itself in fact it is the network capacity utilization at a certain time, like in the evening hour when everybody wants to watch a movie or TV-show. And this is where the network architecture comes in. Different architectures have specific benefits for miscellaneous demands. Choosing the adequate architecture needs knowhow of these systems.

1.4.6.1 Passive Wavelength Division Multiplexer (PWDM) Over Fiber

The Over Fiber solution is one of the simply one compared to the Reconfigurable Optical Add-Drop Multiplexer (ROADM). This architecture contains a video head end office

(VHO), video serving office (VSO) and a hub. Each of these nodes is connected through a fiber cable. The maximum distance between them is about 10-50 km [23]. The VHO is main data warehouse where all video streams come from. Usually the VHO are directly connected to the TV-stations. The hub is responsible for redirecting the data stream to the demanded VSO. And at the end of this chain the VSO are transmitting the data streams to their district or small town. Nowadays these types of architectures are rarely used, because of their limited possibilities and stiff administration. To expand such systems, it needs more fiber cables. In fact this is very cost intensive and often not cost effective.

1.4.6.2 ROADM

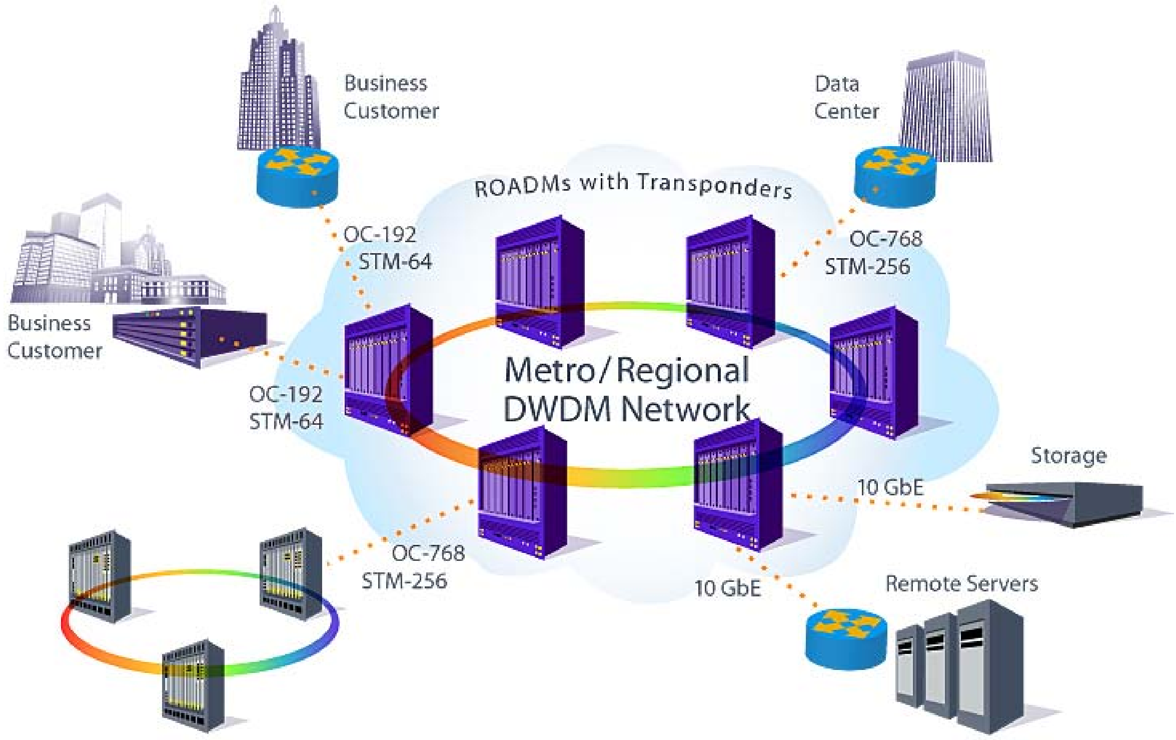

Figure 1.9: ROADM architecture [20]

Other than the PWDM the ROADM architecture is much more flexible [23]. The main architecture of ROAMD is quite the same as the PWDM. The only and deciding difference between these systems are the multiplexers, which are installed on every VSO. Through these multiplexers, it is possible to control the data stream of all VSO in the system. With this ability it is possible to redirect data stream to the VSOs with the most capacity left without leaving your working place. Figure 1.9 shows a possible architecture for ROADMs.

1.4.6.3 Comparison

Deciding between the different architectures to choose may be easier by looking on some key fact of costs. In the following table its shows how the costs of these architectures are changing by raising the demand of Gigabit Ethernet (GbE). Figure 1.10 shows that the ROADM architecture is more effective in large GbE demand, keeping the cost nearly constantly.

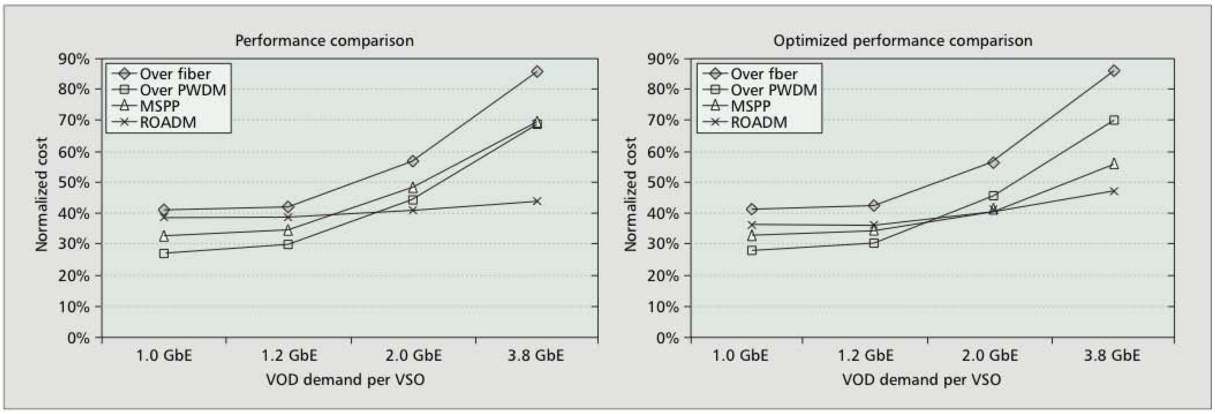

Figure 1.10: Architecture Costs [23]

1.5 Summary

In conclusion it can be said that IPTV has a big potential in the world market. The steady growth of subscribers, investments in this technology and quantity of companies shows that it is not a short term trend. Main drivers of this technology are the industry and the steady growth of bandwidth capacity. Although the cable industry still has the major market share, IPTV is getting more and more valuable in the world market. IPTV can nowadays offer the same Quality of Experience as cable does because of the Quality of Service assurance from the service providers. Furthermore IPTV offers a big potential in creating new capabilities in TV-experience like interacting with the TV or help shifting “computer activity” to the TV.

Bibliography

[1] Akira Takahashi, David Hands, Vincent Barriac: *Standardization Activities in the ITU for a QoE Assessment of IPTV*, IEEE Communications Magazine, February 2008.

[2] Bart De Vleeschauwer, Wim Van de Meessche, Pieter Simoens, Filip De Turck, Bart Dhoedt, Piet Demeester: *On the Enhancement of QoE for IPTV Services through Knowledge Plane Deployment*, BroadBand Europe, December 2006.

[3] Charles Waltner: Telcos Tuning In to IPTV; `http://newsroom.cisco.com/dlls/2007/ts_121107.html`, December, 2007.

[4] Christoph Baumgartner: Die schöne neue Welt braucht Bandbreite; `http://www.computerwelt.at/detailArticle.asp?a=123422&n=25`, August, 2009.

[5] Comparis: TV Angebote im Vergleich; `http://www.comparis.ch/telecom/tvvergleich.aspx`, April, 2009.

[6] Eric Linask: IPTV Market to Slow, but still Grow; `http://iptv.tmcnet.com/topics/iptv/articles/57220-iptv-market-slow-but-still-grow.htm`, June, 2009.

[7] Goldmedia: Telekommunikations- und Pay-TV-Dienste treiben Kabelindustrie an; `http://www.goldmedia.com/aktuelles/suche/info/article/telekommunikations-und-pay-tv-dienste-treiben-kabelindustrie-an.html`, May, 2010.

[8] Heise-Online: Telekom vermarktet IPTV ohne Internetzugang; `http://www.heise.de/newsticker/meldung/Telekom-vermarktet-IPTV-ohne-Internetzugang-860527.html`, November, 2009.

[9] IPTV-Anbieter: IPTV Einführung; `http://www.iptv-anbieter.info/was-ist-iptv.html`, May, 2010.

[10] IPTV-Anbieter: Alice TV - Der große Praxis Testbericht; `http://www.iptv-anbieter.info/testberichte/alice-hometv/alice-iptv-test.html`, May, 2010.

[11] Irving E. Fang: *A history of mass communication: Six information revolutions*, Focal Press, New York, 1997.

[12] ITU-T Study Group 2: *Teletraffic Engineering Handbook*, ITU-D, Juni 2006. http://www.com.dtu.dk/teletraffic/handbook/telenook.pdf

[13] Jochen Wieloch: 15 Millionen IPTV Nutzer in Deutschland; http://news.magnus.de/internet/artikel/iptv-15-millionen-nutzer-in-deutschland-zahlungsbereitschaft-fuer-hd-inhalte.html, May, 2010.

[14] Joan Van Tassel: Piracy or DRM: Either way, there's a high price to pay; http://www.hollywoodreporter.com/hr/search/article_display.jsp?vnu_content_id=1000930093, May, 2005.

[15] K. C. Jones IPTV Revenue, Competition heating Up; http://www.informationweek.com/news/internet/showArticle.jhtml?articleID=201000471, January, 2007.

[16] Ken Kerpez, Dave Waring, George Lapiotis, J. Bryan Lyles, Ravi Vaidyanathan: *IPTV Service Assurance*, IEEE Communications Magazine, September 2006.

[17] Klein Report: Weiterhin rote Zahlen für Bluewin TV; http://www.kleinreport.ch/meld.phtml?id=41028, June, 2007.

[18] M. Cotton, L. Vegoda, D. Meyer: *IANA Guidelines for IPv4 Multicast Address Assignments*, Internet Engineering Task Force (IETF), March 2010. http://tools.ietf.org/html/rfc5771

[19] Michael Peng: IPTV in China : Market Prospects; http://www.heavyreading.com/details.asp?sku_id=1259&skuitem_itemid=960, September, 2006.

[20] Mintera: Mintera Products and Solutions ROADM Capability; http://www.mintera.com/productpages/productsroadm.htm, May, 2010.

[21] Mojca Volk, Joze Guna, Andrej Kos, and Janez Bester, University of Ljubljana: *Quality-Assured Provisioning of IPTV Services within the NGN Environment*, IEEE Communications Magazine, May 2008.

[22] RedChip Blog: New Gps Product has ZTSN on a Path to Growth; http://blog.redchip.com/index.php/2010/headline/new-gps-product-has-zstn-on-a-path-to-growth/, January, 2010.

[23] Sunan Han, Sam Lisle, Greg Nehib: IPTV Transport Architecture Alternatives and Economic Consider- ations; IEEE Communications Magazine, February 2008.

[24] Swisscable: Swisscable > Angebote TV-Kabelnetze > Fernsehen > Analoges Fernsehen; http://www.swisscable.ch/article/article.php3?art=23, April, 2010.

[25] Swisscom: Swisscom Fixnet lanciert Bluewin TV im Jahr 2006; http://www.swisscom.com/GHQ/content/Media/Medienmitteilungen/2005/20050526_03_bluewin_tv.htm, May, 2009.

[26] Swisscom: Swisscom Q1: Umsatz leicht höher - Fastweb belastet Gewinn; http://www.bluewin.ch/de/index.php/1651,273106/Swisscom_Q1__Umsatz_leicht_hoeher_-_Fastweb_belastet_Gewinn_\%28Zus\%29/de/news/boerse/kurse/schlagzeilen/, May, 2010.

[27] Thomas Riegler: *IPTV - Kabelfernsehen aus dem Internet*, Verlag für Technik und Handwerk, Baden-Baden, 2008.

[28] Wes Simpson, Howard Greenfield: *IPTV and Internet Video: Expanding the Reach of Television Broadcasting, Second Edition*, Focal Press, New York, 2009.

[29] Wikipedia: DRM; `http://de.wikipedia.org/wiki/Digitale_Rechteverwaltung`, 2009.

Chapter 2

Economic Challenges of Standardization

Manuela Züger, Simon Poltier, Andreas Volkart

This seminar paper discusses the implications of technical standards for competition in the context of the Internet. The process for the definition of Internet standards is first described and commented. A simple model and some examples are then used to show strategic interaction with respect to standard-related decisions, illustrating the influence of network externalities and competitors' expectations on strategies. Finally, case studies on the Browser War and on HTML5 / Flash as Web standards are used to examine and extend the insights of the first two parts, emphasizing in particular the influence of externalities.

Contents

2.1 Introduction and Problem Statement

Standards are of particular importance in technical environments. They help to create a solid technical infrastructure, which will serve as a foundation for innovation-led growth. [1] Especially the fast growth and development of the Internet would not have been possible without having well-defined and generally accepted standards.

On the economic side companies can benefit from entering possibly large markets by making their product compatible to a certain standard. By doing so technological risks can be reduced, because the monetary exposure of developing a specification is far higher than producing a product conform to a well-established standard. Market risks can be reduced as well on the supplier side, because the company, which follows a standard is less vulnerable for changes of market variables. However one can observe that there are firms, which are economically successful in making proprietary specifications and not following any standard at all. This phenomenon will be encountered in the case study of Section 2.4 once again.

Customers can potentially profit from lower prices and better products, because there is often a considerably high competition within a standard. Moreover one can benefit from a wide variety of complementary goods by choosing a product applying to well-established standards.

This paper focuses on economic challenges of standardization. It describes especially how to compete with other firms using or promoting own or foreign standards. Two case studies on web browsers and web videos will help to understand and evaluate the proposed theories. But first of all the following sections will look at standards in general as well as Internet Standards.

2.1.1 Definition

There are mainly two kinds of standards: proprietary standards and open standards. They strongly differ in the process they have been created.

Open standards are developed by an official gremium in a well-defined manner. Concerning open standards the European Union defined the following four key points [2]:

- The standard is adopted and will be maintained by a not-for-profit organisation, and its ongoing development occurs on the basis of an open decision-making procedure available to all interested parties (consensus or majority decision etc.).
- The standard has been published and the standard specification document is available either freely or at a nominal charge. It must be permissible to all to copy, distribute and use it for no fee or at a nominal fee.
- The intellectual property - i.e. patents possibly present - of (parts of) the standard is made irrevocably available on a royalty-free basis.

- There are no constraints on the re-use of the standard.

Many countries and organisations have very similar definitions. They all emphasize the openness of the specification process and the unrestricted and free usage of the standard itself, which means that no costs shall arise when using an open standard. A good example for open standards are members of the Internet protocol family like IP, TCP, HTTP or SMTP, which have been specified and released by the Internet Engineering Task Force (IETF) as Internet standards.

In contrast to this, there are so-called proprietary standards promoted by profit-driven organizations. They do not necessarily fit the four key points of open standards listed above. It is common that these standards do not arise from an open process and that specification documents are not available for free or not available at all. Furthermore some patents may be part of the standard, which often means that the usage is restricted. A famous example is the VHS standard which has been released by JVC in 1976.

Proprietary standards may gain a dominant position in the market. These specifications are called de facto standards, because of their widespread usage and acceptance. Nonetheless is important to notice that they don't necessarily need to be officially certified or developed in an open process. An example is the Microsoft Word document format. Because of the dominant market position of Microsoft Word in office environments it is supported by almost every manufacturer of word processing software via import or export filters. It is often the case that such de facto standards get officially certified after they have reached this dominant position. An example for this proceeding is the proprietary Portable Document Format (PDF) published by Adobe, which became an ISO-norm in 2001. [3]

2.1.2 Internet Standards

Concerning Internet standards one organization is of particular importance: the Internet Engineering Task Force (IETF). The main goal of the IETF is to "'make the Internet work better"'. They try to achieve this goal by releasing different types of formal and informal technical documents such as Best Current Practices (BCP), Internet Drafts or Internet Standards. The IETF ia an open organization, in which everyone who is interested can participate without being a member of this Task Force. Strictly-speaking there is no real membership in the IETF. A quotation of David Clark, the former chief of the Internet Architecture Board (IAB) about the workflow in the Internet Engineering Task force shows nicely the philosophy of this organization: "'We reject kings, presidents and voting. We believe in rough consensus and running code."' [4] Internet Standards are developed by several working groups which use mailing lists and meetings for discussion and coordination. As the IETF is an internationally acknowledged institution, their Internet standards are used and accepted worldwide.

The Internet Engineering Task Force tries to provide open standards [4]. So every participant in the developing process is responsible that the contribution brought in is free of property rights of third parties. If there are such rights they must be opened to everyone to guarantee the openness of the standards.

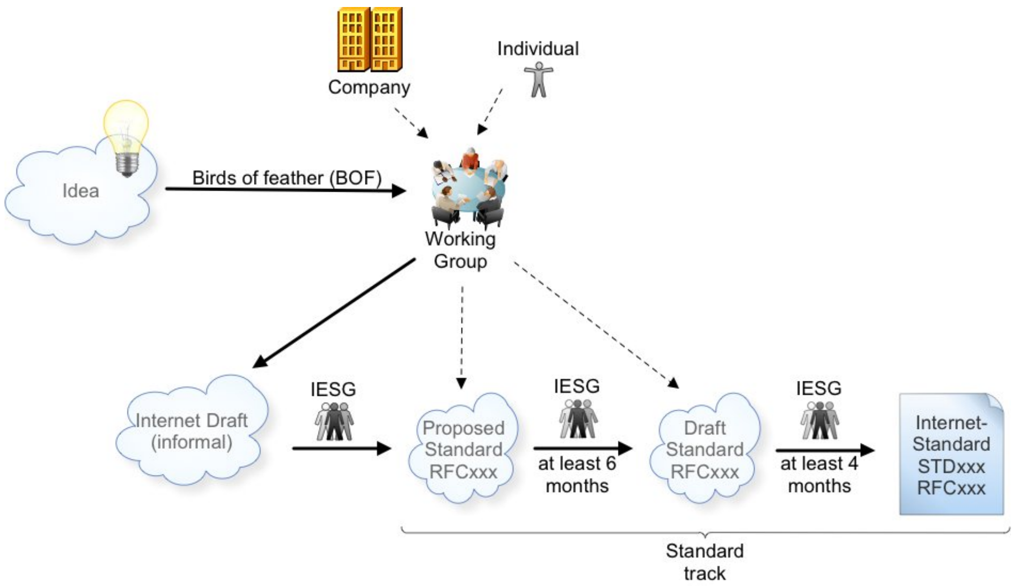

Figure 2.1: Standardization Process of the IETF

The IETF uses a well-defined process to get from a raw specification to a final Internet Standard. The exact process is described in the document RFC 2026.[5]. First of all Birds of feather (BOF) meeting is taking place. The goal of this informal meeting is to find out whether the proposed specification has enough potential and gains enough community feedback to launch a IETF working group. If this is the case, a working group is formed. In the following this working group will elaborate so called Internet Drafts. These are informal documents published on the IETF website to gain as much as possible community review.

The next step towards a Internet Standard is taken by the Internet Engineering Steering Group (IESG), which is responsible for the management of the IETF. A specification that is fairly stable, has gained enough community review and is considered valuable is moved to the so-called "'standards track"'. This is a series of three maturity levels, through which every specification meant to be an Internet Standard must go. The three maturity levels are called "'proposed standard"', "'draft standard"' and "'Internet standard"'. Every specification entering the "'standards track"' is given a RFC-Number and is published as a RFC-document on the website of the IETF. RFC means Request For Comments and shows the intent of such a document. The community should comment on these specifications. A proposed standard remains at this level for at least 6 months. In this time at least 2 implementations with different code bases should be made to test the standard and gain further experience. Those are mandatory requirements to advance in the standards track to the next level.

According to the requirements mentioned above the Internet Engineering Steering Group can decide to make the proposed standard a draft standard and publish it again online. A draft standard is already quite stable and only minor changes should be applied during this period. The draft period lasts at least another 4 month until the IESG can finally take the decision to make the draft standard a Internet Standard.

During the whole standards track the specification keeps the same RFC-Number. An Internet Standard has an additional STD-number.

If there are any conflicts within the IETF groups during the development of an Internet Standard, a well-defined and open discussion takes place. If the conflict cannot be resolved this way, the IAB (Internet Architecture Board) as the mother organization of the IETF has the final say. it is also important to notice that although there are several objective criteria as mentioned above, there is no clear algorithm that defines the rules for elevation on the standards track. So the technical competence of the IESG is of particular importance for the whole process.

Due to the openness and the involvement of many parties in open organizations as the IETF, it takes quite some time until the standard is finally released. So the standardization process often cannot cope with the rapidly changing technology. This fact often meets with criticism as in [6]. A proprietary standard of a particular firm may adopt the current technology faster than the open standardization process. This problem can be viewed as a trade-off between the technical maturity and the time a standard can be developed. One might argue that for the internet it is more important to have sophisticated standards because they preserve longer than other technical artefacts as for example file formats. In this sense these open processes are the way to go. On the other hand there are also examples in which this slowness is not particularly useful for the Internet as an open medium. For example during the development of HTML5, which will provide a general support for Internet videos (video tag), the proprietary Flash video standard of Adobe has already gained a dominant position. This example is seen once again in Section 2.4.

W3C (World Wide Web Consortium) is another, well-known gremium, which develops open specifications for the Internet. It concentrates on content-related techniques such as HTML or XML. It is important to notice that the W3C does not publish Internet Standards but W3C-Recommendations. Some of these recommendations, in particular HTML, became de facto standards in the Internet or served as a basis for ISO-standards. However the World Wide Web Consortium is not internationally acknowledged and therefore strictly-speaking not allowed to define Internet Standards.

2.1.3 Motivation and Problem Statement

In the Internet and IT world, standards are the object of intense conflicts between companies, and the victory of a standard over another can have tremendous consequences on these markets and the technology they end up using.

The goal of this seminar paper is therefore to explore and acquire a first understanding of how standards interact with the competitive environment, in general economic terms and then as applied to the Internet. Besides it will try to find out why firms adopt certain objectives with respect to standards, and what strategies they use in order to achieve these objectives.

2.2 Standards and Competition: Economic Analysis

In this section, the main characteristics of competition with respect to standards are discussed, starting with the general economic features of standards. A model of the strategic behavior of competitors is then developed, with special care given to the importance of standards for that behavior.

2.2.1 Network Economics

Standards, especially communications standards as in the case of the internet, are what economists call networked goods, since users tend to benefit when other users join the standard.

The additional benefits users of a networked good obtain as the number of users grows is called a network externality. There are two main types of those: direct and indirect network externalities [11].

Direct network externalities are the benefits the user gets from being able to access new users through his access to a network, for example the ability to access a new server through the Internet (assuming that the server was previously using a different, non-standard set of protocols). The link to the network itself is worth more to the user when more users join the network.

Indirect externalities appear as side effects of a standard becoming widespread: markets for complementary products that are compatible with the standard emerge, and these products are more widely available and affordable as a result [11].

In the case of the Internet, know-how is a good example: it will be easy to find experts on the different protocols in the Internet stack, whereas a business that develops its own protocols will have to train experts in-house, which will be more time-consuming and expensive.

Users are not the only ones to benefit from networked goods. Those who offer standardized goods or services benefit from different types of economies allowed by networked goods. Economies of scale occur because more users can be served using the same technology, e.g. users of mobile phones in GSM-only countries can be served with lower costs than users in countries where several cellular standards coexist.

Economies of scope also appear, since different goods or services can be offered using the standard technology; a simple example of this is the existence of one electric plug standard (at least on a national level) to power most types of electric devices, thereby lowering manufacturing costs for these devices.

2.2.2 Impact of Standards on Demand

For a communications related service, the presence of a standard contributes greatly to reducing interoperability costs, e.g. the costs attributed to a link. However, switching to the standard implies costs at the node, without certainty that the other users will agree to switch ([11], p.3).

That is why, If interoperability is important, the users will often wait to invest in a new technology until one standard emerges as the dominant one ([7], p.3), which in turn makes it more difficult for willing users to be early adopters.

[8] discusses the inertia effects caused by networks: users will enter a communication network (e.g. start using a standard) if the standard reaches a critical mass, and this critical mass is composed of the right elements.

In the Internet case, examples of "'the right elements"' would be the most needed servers/services from the user point of view, and the most valuable, e.g. higher paying, users from the server point of view. This results in an upside-down U-shaped demand curve, where depending on the existing amount of users, the demand will stabilise at a different value.

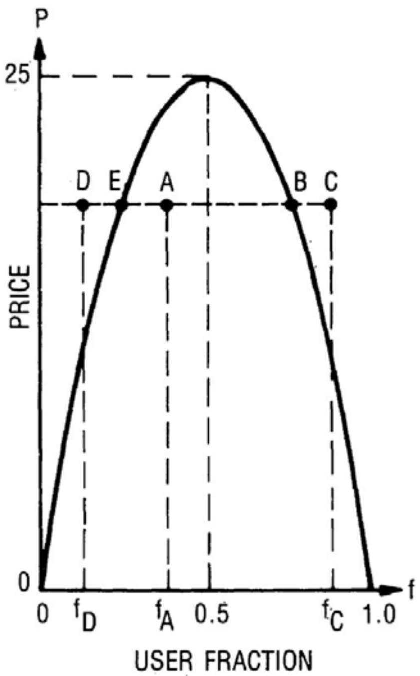

Figure 2.2: U-Shaped Demand Curve

Figure 2.2 from [8] shows the equilibrium price as a function of the fraction of users who are part of a network in a given population. For the price emphasized here, critical mass corresponds to point E, and at any point under the curve with the same price (such as A), the network is worth more to the user than the fixed price, so the number of users will grow (up to point B in the case of point A).

On the left side of the curve (e.g. at point D), the price is too high and users will leave the network one after the other, moving the equilibrium back to zero users with a price of zero.

2.2.3 Role of Standards Organisations

Standards organisations are different depending on the field and country. Different competitors or actors in the system may not have the same access to a Standard Setting Organisation (SSO), because of lack of financial or technical resources, geographical location, or other reasons.

Mattli and Büthe [10] propose that a high degree of hierarchy at the national level helps to bring more interested parties in the discussion (by offering more, smaller venues for discussing standards). A high degree of integration and regulation of standards also helps to integrate with international standards setting organisations, as opposed to a non-regulated, competitive environment where private Standard Development Organisations (SDOs) have to keep their drafts standards secret and sell copies of final standards at a high price to remain viable, which results in duplication of standards and a lack of unique standards at the national level, rendering integration difficult.

This variation in efficiency will also have an impact on the time-to-market penalty incurred by using an official SSO; a firm will tend to bypass SSOs if it expects its standard to take a long time to be certified if submitted.

2.2.4 Competition and Standards

Because of the behaviour of demand, a provider cannot sell his network service for a sustainable price unless he reaches a critical mass of users. In order to attain critical mass, a demand-encouraging price policy will be required. Finding a way to select the users who are the most willing to pay, so as to have a stable userbase when one starts to increase the price, will also be necessary.

One of the ways this can be achieved is, when possible, by using price discrimination. A critical mass can thus be achieved at a positive price as more and more users adopt the standard, increasing its value [8].

Additionally, positive indirect externalities can be influenced by encouraging the supply of complementary goods and services compatible with the standard. For a competitor who is not the first to enter a market, pre-announcing one's product may help to reduce the first mover's lead [7].

If a competitor is a dominant player, it will have the power to set a proprietary standard with favorable characteristics for itself. Depending on the quality and accessibility of the official standards-setting process, this may be an optimal solution.

2.2.4.1 Model of Standard-related Competition

As seen in [7], the important strategic questions when it comes to standards in a competitive marketplace are

- Is it better for competition to happen within or between standards ? (from some or all competitors' perspective)
- Do one or more competitors want to impose their standard (for reasons of costs / technical advantages)

A simple 2-player game-theory model with 2 different standards is used. Actions for each player are to use standard 1 or 2, as shown in Table 2.1.

Table 2.1: The simple two-player game [7]

		Firm II	
		Standard 1	Standard 2
Firm I	Standard 1	a_{11}, b_{11}	a_{12}, b_{12}
	Standard 2	a_{21}, b_{21}	a_{22}, b_{22}

Payoffs in Table 2.1 will depend on the elements explained above, e.g.: if critical mass is hard to achieve, i.e. network externalities for users are strong, players will want to compete within standards rather than between, in order to feed off each other's critical mass of users. The availability and credibility of SSOs will make it easier to agree on a common standard as well.

Those factors will result in distinct strategic situations. Competitors could have the following preference combinations:

1. both want to use a given standard, e.g. a11 and b11 are the highest payoff for both firms. If $a_{22} > a_{12}$ and a_{21}, $b_{22} > b_{12}$ and b_{21}, there still is a possible equilibrium in (Standard 2, Standard 2) since if one firm chooses 2, the other firm is better off choosing 2 than 1.
2. both want to use one standard together rather than use a different standard, but disagree on what that standard should be; e.g. $a_{11} > a_{22} > a_{12}$ and a_{21}, $b_{22} > b_{11} > b_{12}$ and b_{21}
3. each wants to use his own standard without the other having access to it, e.g. a_{12} is the best for firm I, b_{12} for II
4. One wants to use a common standard, the other one wants to keep his standard to himself: b_{11} or b_{22} is greater than b_{12} and b_{21}, but for firm I, a_{12} is the largest possible payoff.

In a second stage, uncertainty is introduced in the model using a simpler version of the approach in [9]: firms are of different types (defined by their situations with respect to standards, their status in the market, their clients).

Different types have different optimal strategies, and firms do not know which type their competitor is.

Concretely, this means that there are 4 types of firms in the market, which are referred to as A, B, C, and D. Types A and D prefers to share standards, types B and C prefer not to; types A and B prefer standard 1 whereas types c and D prefer standard 2.

Each of these types has a different set of payoffs for each of the 4 outcomes in table 1, and each represents a certain fraction (F(A), F(B), etc) of all possible competitors. A given player has to take this distribution of types into account when devising his/her optimal strategy.

An example is the situation where a player of type A must devise a strategy in a population of A, C and D types. The different possible games for this player look like the following:

Table 2.2: A player of type A playing against a player of type D

		Firm II (D)	
		Standard 1	Standard 2
Firm I (A)	Standard 1	4, 2	1, 1
	Standard 2	1, 1	2, 4

Table 2.3: Players of type A playing against one another

		Firm II (A)	
		Standard 1	Standard 2
Firm I (A)	Standard 1	4, 4	1, 1
	Standard 2	1, 1	2, 2

Table 2.4: A player of type A playing against a player of type C

		Firm II (C)	
		Standard 1	Standard 2
Firm I (A)	Standard 1	4, 1	1, 4
	Standard 2	1, 1	2, 1

If F(D) is sufficiently high, D-types will always play Standard 2 knowing that most of the time, they will face a D-type. Being faced with this majority, A-types will be better off playing Standard 2 as well, to get the benefits of cooperating, rather than trying to benefit from the few times they are matched with an A-type.

This strategy will also prevent them from being exploited by C-types, who would like their opponent to choose standard 1.

Even in the case where both firms would like to use standard 1, they will not do so because of the risk that they are playing with an opponent whose preferences are different.

The fractions of each type can also represent the factors that influence firms in their evaluation of standards; for example, if standard 2 was technically complex to implement, firms with fewer technical resources (i.e. engineers, etc) would be C- or D-types.

To attempt to correct this inefficiency, firms will use their knowledge about their competitors and try to influence their competitor's knowledge.

To represent this aspect, the game is extended with an initial phase where firms declare their type, before choosing a strategy, as also seen in [9]. Firms do not have to be truthful in that initial phase.

For the numerical example above, this means that players of type A will be able to say that they are type A; D-types are indifferent in revealing their type since they are in the majority. What is important for A-types is that they can now change their choice to Standard 1 if their opponent declares type A: cooperation is made possible through the exchange of information.

The remaining issue is C-types: they have an interest in A-types choosing standard 1, so that they can be the sole firm in standard 2. They will therefore declare their type as A, in order to trick A in choosing standard 1.

Depending on the amount of C types in relation to A types, this will reduce or eliminate the positive returns of collaboration for firms of type A.

2.2.4.2 Further Strategies

Finally, firms can use more sophisticated strategies to influence their competitors. Among those, commitments are decisions that reduce one's payoff in case of switching to the competing standard, reducing the competitor's hope that one will adopt their standard.

On the opposite, concessions are decisions that increase the competitor's payoff if they switch to one's standard (e.g. lowering of license fees) or decisions that show a willingness to switch to the competing standard (e.g. making one's products partially compatible with that standard) [7].

2.3 Case Study: The Browser War

The Browser War is probably the best-chronicled standards competition in recent history.

It is an interesting example for standard diffusion and network effects. Furthermore it can be shown what role Internet standards play in this battle. It is a competition between standards because the companies behind the browsers want to lock in users by building a new de facto standard for example by extending HTML through proprietary tags which provide additional functionality.

2.3.1 History of the Browser War

The Browser War can be divided in three acts. The competition battle started with the first act in 1994 when Netscape entered the market and put an end to the absolute monopoly of Mosaic which was the first web browser available. In 1995 the second act began when Microsft launched the distribution of its browser Internet Explorer (IE) and forced Netscape out of the market. Later, additional suppliers of web browsers entered the market and the competition increased. Today, the third act is still undecided.

2.3.1.1 Act 1: NCSA Mosaic vs. Netscape Navigator

Mosaic was the first graphical web browser and became market leader because it was the only web browser available at all. When Netscape entered the market with its browser Netscape Navigator, Mosaic quickly lost its users. Netscape's success was based mainly on its new price strategy, giving the browser away for free to private users and price discriminating towards business users, who are not as price sensitive as private users and presumably get added values for what they pay (e.g. support). Netscape's long term goal was not to give its browser away for free to all users. By making sure some of them are willing to pay, it ensures that part of its user base will accept to pay for the browser. The other factor of Netscape's success was that its browser was more performing than Mosaic. For example it was ten times faster in loading pages and it was enriched of features that allowed web publishers to create more attractive and performing web pages. When Netscape became market leader it exploited its market position and charged end users as well [12].

2.3.1.2 Act 2: Netscape Navigator vs. Microsoft Internet Explorer

In 1995 Netscape had a 90% installed user base for its browser. Internet Explorer 1.0 and 2.0 were licensed versions of NCSA Mosaic. They had a cost like Netscape Navigator. After Bill Gates famous "'Pearl Harbor"' speech, where he proclaimed that Microsoft would become a leading internet player, the strategy of Microsoft changed. From now on IE was distributed for free. Netscape had to react quickly and gave away its browser for free as well. In 1997 Netscape was losing market share and money at an alarming rate. Netscape had no other income stream and was taken over by AOL in 1999 [13].

2.3.1.3 Act 3: Microsoft Internet Explorer vs. Mozilla Firefox

The third act of the Browser War takes place between the no-profit organization Mozilla, whose objective is to promote the open source browser Firefox, and Microsoft, who distributes the browser Internet Explorer. Mozilla Firefox is a multi platform open source web browser that has been developed in accordance with the World Wide Web Consortium (W3C).

Mozilla 1.0 was initially a superior product in term of security regards to IE 6.0. But, no price strategy advantage has been played by Mozilla Foundation, since Firefox is released free of charge for private users as IE. Also, IE resolved very soon its security problem, since in early 2006, Microsoft released a product not inferior to Mozilla Firefox 2.0.

Nevertheless, IE is still losing market shares. Ventura et al. assume this is because IE is still paying negative network externalities due to being the most popular web browser. Indeed, internet attacks are mostly developed for popular applications. This implicates that some users prefer to switch to a less popular browser like Firefox in order to escape potential attacks. [12].

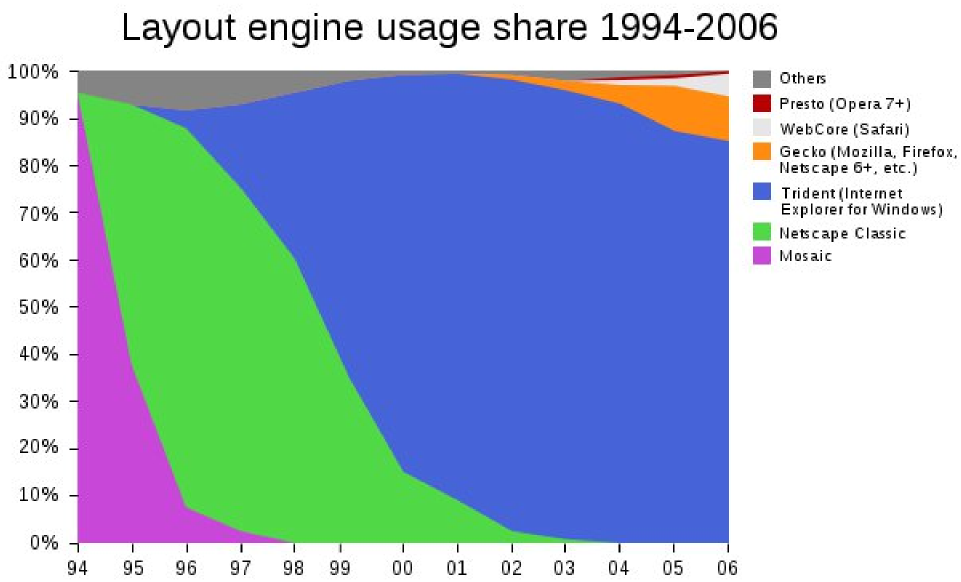

Figure 2.3: Layout Engine Usage Share 1994 - 2006 [18]

2.3.2 Analyses of the Second Act of the Browser War

The second act of the browser war, namely the one between Netscape Navigator and Microsoft Internet Explorer will be analyzed in matters of web standard implementation and economic strategies. It serves as an example of a standards competition between two standards. In this case the players would be of types B and C in the 2-player game-theory model.

2.3.2.1 Browsers as De Facto Standards

The browser included new non-official HTML tags that extend the possibilities of web design. This creates a parallel de facto standard which helps locking customers in.

Examples of Netscape's de facto standard:

- `<frameset>`

  HTML frames allow authors to present documents in multiple views, which may be independent windows or sub windows. Multiple views offer designers an option to keep certain information on the screen, while other views are scrolled or replaced [19].

The frames where established by Netscape and are later considered as a normal part of web page design. www.digitec.ch is still using this technology.

- `<blink>`

  This tag makes the text in it blink.

Examples of Internet Explorer's de facto standard:

- ActiveX controls:

  These are small program building blocks which can be embedded in HTML content. They can serve to create distributed applications that work over the internet through web browsers. Examples include customized applications for gathering data, viewing certain kinds of files and displaying animation. They are used for example by Windows Live Hotmail [20].

- `<comment>`

  This tag can be used to add a comment to a HTML file.

- `<bgsound>`

  This tag adds some background music to a web page [14].

2.3.2.2 Economic Analyses of Microsoft's Success

It is widely recognised that IE 3.0 and 4.0 were technologically inferior to Netscape Navigator 2.0 and both companies followed the same price strategy by giving away the browser for free.

Additionally the start-up is very crucial and important. Microsoft could only win, because it had good starting conditions.

There were basically two aspects which helped Microsoft in the start-up:

- Internet users do not like to switch from the first browser they come in contact with. This means that if a market entrant can ensure that his browser is the first which users come in contact with, this users will probably not switch to the concurrence anymore.
- At the time of the browser war, the internet grew at an enormous rate. In fact, it doubled in size each year. This implicates that a market entrant would be able to dominate the whole market soon, if he only gets the majority of the future internet users.

This provided ample scope for a late entrant [13].

The Microsoft strategy:

Most users were either acquiring their browser in a hardware or software bundle, or else through their Internet Service Provider (ISP). Microsoft bundled its browser software to the Windows operating system to use the Windows installed user base to boost the Internet Explorer adoption and accelerating related positive network externalities. Further Microsoft used its market power over the PC retail and the ISP markets to ensure that IE and not Netscape Navigator was the first browser users would come into contact with.

This strategy brought great success to Microsoft but also a conflict with the US antitrust law [12].

Microsoft does not have to directly earn money through Internet Explorer (even in the long term), since it only wants to have a complementary product (IE) that is a standard and is exclusive to its ecosystem. In particular, it is not available for Linux; that means once most websites rely on IE, it gives the Windows platform a competitive advantage. Microsoft succeeded in locking business customers in by making IE a relatively compelling platform for businesses to develop their applications. The business customers had the high willingness to pay required by [8], and they created the critical mass that would keep using IE even if it became less attractive to an extent. The effects of that strategy can still be observed today in certain legacy applications, even though Microsoft does not pursue that strategy anymore.

It is interesting to note comments made by Jim Clark, Netscape CEO, on this question;

"'At some level, (open) standards certainly play a role, but the real issue is whether there is a set of people, a set of very powerful companies, out there who don't play the standards game. For the standards game to work, everyone has to play it, everyone has to acknowledge it's a game. Companies such as Microsoft aren't going to sit around and wait for some standards body to tell them 'You can do this'. If your philosophy is to adhere to the standards, the guy who just does the de facto thing that serves the market need instantly has got an advantage [13]."'

This suggests that companies will seek to capture a de facto standard whenever this grants a competitive advantage. In the case of HTML this involved the privatisation of a previously open and common standard. This is not only true for Microsoft but for all competing for control of the Internet. It was Netscape, not Microsoft, who first used this tactic - adding its own proprietary extensions to HTML - in order to win a standards battle. Research suggests that there are two important conditions in Internet competitions which causes companies to build a de facto standard. These are a high degree of market uncertainty and that no company has previously developed a core technology capable of becoming a de facto standard [13].

2.3.3 Current Situation

Today, IE is the most used web browser followed by Firefox. There are also other browsers available like Opera, Safari and Chrome. The suppliers of web browsers still try to build a new de facto standard and therefore do not consequently hold to the web standards. The Acid tests are an initiative against these efforts.

2.3.3.1 Acid Tests

"'Eines der grössten Probleme, die Webentwickler in der Implementierungsphase bereitet werden, ist die inkonsistente Darstellung der Seiten in verschiedenen Browsern"' [15]. Because of that problem the Acid tests have been developed by the Web Standards Project. The Web Standard Project declare themselves as a "'grassroot coalition fighting for standards which ensure simple, affordable access to web technologies for all"'. The most recent Acid Test is ACID3 which tests W3C recommended standards like HTML, DOM, DOM2, CSS, EcmaScript, ect [16]. It consists of 100 subtests and is considered as "'passed"' when all subtest are passed. The results of the currently popular browsers are as following:

Table 2.5: Results of ACID3

Browser	ACID3 result
Internet Explorer 8.0	20 / 100
Internet Explorer 9.0	55 / 100
Firefox 3.6	94 / 100
Opera 10.0	100 / 100
Safari 4.0	100 / 100
Chrome 3.0	100 / 100

2.3.3.2 Browser Market Shares

IE is the market leader and does not pass the Acid Test which shows that it is not absolutely necessary for a browser to support all web standards to be successful. Because of the large market share of IE, developers have to adapt their web sites to be properly viewed in IE.

2.4 Case Study: Adobe Flash vs. HTML5

Today there is a strong need for multimedia contents directly within the browsers. Especially web videos became very popular in the last few years. But also other technoligies like animation, vector graphics or drag and drop are of particular importance in the Internet environment. The trend goes towards the integration of local applications into the browser. Those applications are called Rich Internet Applications.

Two basic technological platforms have the potential to support all these needs. On one hand the proprietary Flash platform and on the other hand the open HTML5 technology, which is still in the development stage. In terms of standards the video functionality of Flash and HTML5 is an interesting topic to look at.

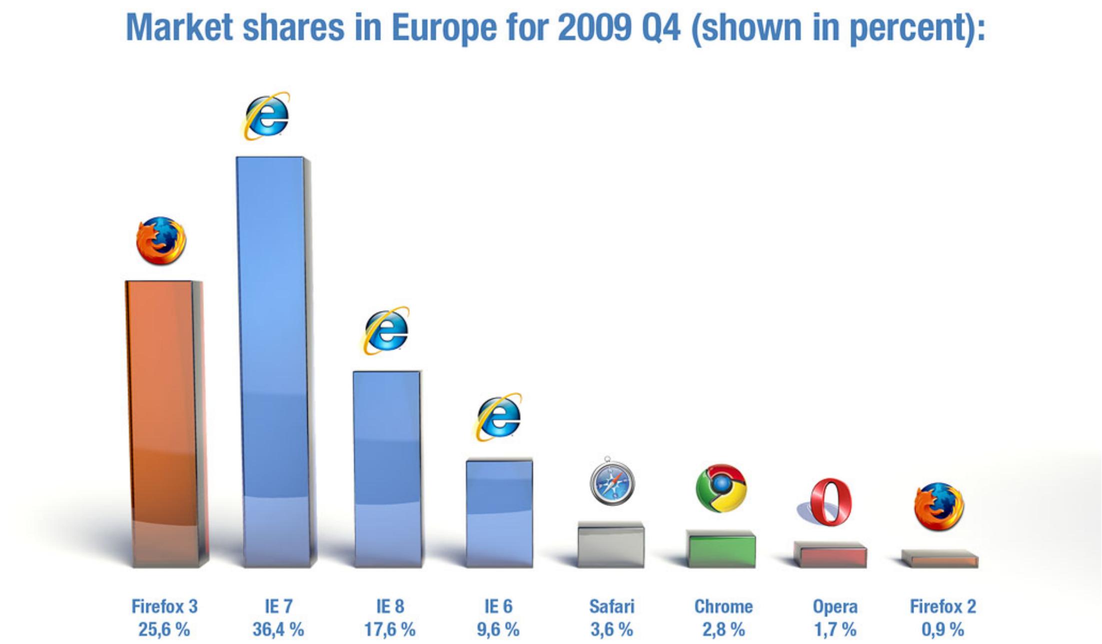

Figure 2.4: Web Browser Market Shares in Europe for 2009 Q4 [21]

2.4.1 Adobe Flash

Flash is an interactive multimedia platform developed by Adobe. It can be integrated into all modern web browsers by a Flash Player plugin. According to Adobe 99% of all web users have this plugin installed. [22] Flash gained this immense user base in the last years mainly because of its sophisticated video codec H.264 which is for example used by the most important video platform in the web youtube. One can say that nowadays Flash is a de facto standard for web videos.

Despite of its huge popularity Flash faces some problems. One big problem is, that Flash is hardly supported by any mobile device. For example popular smartphones such as the iPhone don't support Flash. This is because the Flash Player is quite resource-consuming. Also the version for Mac- and Linux operating systems have some performance problems. Another important disadvantage of flash is the to some extent insufficient integration into the HTML code. All Flash applications run in a container within a proprietary `<embed>`-tag, which is not part of the HTML standard. So the rest of the website does not know what is happening in the container and vice versa.

2.4.2 HTML5 `<video>`-Tag

HTML5 is an open standard for Rich Internet Applications developed by W3C. There's not yet a final version but some of its features can already be used, because all important browsers, except Internet Explorer, already support some of them. [23] The `<video>`-tag does not specify a particular video codec.

The basic problem of the HTML5 video functionality is that Internet Explorer does not support it yet. Furthermore there is a need for a free video codec to use within the tag. Free browsers such as Firefox would not pay fees for a proprietary codec such as H.264. So it generally makes sense to use an open codec in an open standard.

2.4.3 Video Codecs

There are mainly three important Video Codecs nowadays, which have to be considered:

- H.264: H.264 is a proprietary format by MPEG LA, which contains patents. Microsoft and Apple are members of MPEG LA but Adobe and Google are not. This codec is widely used in the Internet (e.g youtube). The Flash Player uses it as the standard video format because of its high compression ratio. There are also several hardware-decoders for H.264.
- VP8: VP8 has been developed by On2 an contains patents, too. Google has purchased On2 in 2010. [24] This codec has a similar quality as H.264.
- Ogg Theora: Ogg Theora is an open source video codec. It is based on VP3.2 by On2. The quality is slightly worse than H.264 and VP8. There are concerns that the code could still contain patens of On2. [25]

2.4.4 Current Developments and Outlook

Today the situation is clear. Flash is the market leader in terms of Internet videos. Adobe managed to spread out its product this way mainly because of two reasons:

- Adobe was the first player that could profit from the rapidely growing market for Rich Internet Applications
- Flash can provide all important technologies for Rich Internet Applications including high-quality videos in one stable platform

What is needed to force web designers to use the HTML5 video technology? First of all webdesigners will only use a technology diffrent to Flash if they can be sure that almost every internet user can display their content. For this purpose HTML5 would need a large user base, which it currently not has. The <video>-tag is not yet supported by Internet Explorer, which still has the highest market share. Microsoft recently announced that Internet Exlorer 9 will support the <video>-tag and use the H.264 codec. [26] However the older versions of Internet Explorer will stay in the market for years as can be observed today with the Internet Explorer 6. In this respect the advantage still rests with Flash.

Another reason why Flash will probably stay market leader for quite some time is the unsolved question of the video codec to use within the HTML5 <video>-tag. Internet Explorer, Chrome and Safari promote the H.264 codec as Mozilla and Opera only use the free Ogg Theora. Non-profit organizations like Mozilla will never be able and probably not willing to pay royalty fees for proprietary software containing patents such as H.264. However both Firefox and Opera have considerably large user bases, which makes it impossible for H.264 to become the standard for HTML5 videos. The Ogg Theora codec itself is no real solution because it has inferior compression ration and quality to other technologies. Furthermore there are some concerns about possibly included patents in Ogg Theora.

The only solution to solve this dispute would be a technically sophisticated and completely free video codec. Either MPEP LA (H.264) or On2 owned by Google (VP8) will distribute their codecs for free or a non-profit organization such as Xiph.Org Foundation (Ogg Theora) will develop a technically sufficient format. The latter is rather unlikely because the gap between current open source- and proprietary codecs is too high. In contrast some people consider it likely that Google will make the recently purchased codec VP8 freely available. [25] This company would be able to promote HTML5/VP8 via its leading video platform youtube as well. A fact that supports this theory is the recently installed HTML5 Beta mode for youtube. [27]

Another crucial aspect is the support mobile devices. The battle there is not yet fought. Flash is not yet supported by many mobile internet device. For example the the popular gadgets iPhone and iPad do not support Flash. However this is not a door opener for HTML5/VP8, because in general mobile devices need hardware chips to decode videos. And there H.264 has an advantage because it is already well-supported in hardware. So if Adobe manages to make Flash less resource-consuming it could use the advantage of hardware support to establish this technology on mobile devices as well.

The most likely situation for the near future is that HTML5 and Flash will coexist, with Flash as the de facto standard. If no technically appropriate and free video codec can be found for HTML5, Flash will keep this position for a long time. But if for example Google will release their VP8 codec for free, the `<video>`-tag can become a serious alternative to Flash.

2.4.5 Evaluation and Discussion

As Netscape in the early nineties, Adobe managed to build up a critical mass with its Flash format early. This has several reasons: Flash has been the only complete Rich Internet Application platform at the first stage. In addition the plugin needed to display the content has been available for almost every browser free of costs. It is easy to develop applications for Flash and in contrast to other technologies one can be sure, that the content will be displayed exactly the same way in all browsers using the plugin. So once enough web designers had included the technology in their website, the user got locked in in this technology because many websites using Flash are almost unusable without having a plugin. In the following Adobe used their user base to establish H.264 as a de facto standard for internet videos.

2.5 Summary and Conclusions

It could be shown that standards have a large role to play when it comes to the strategic decisions of businesses. In the case of the Internet, the standards for the communication parts of the technology, as defined by the IETF, are developed and distributed in an open way. On the other hand, the case studies on the browser wars and the standardization of web video show an environment where proprietary standards still matter a lot. Even with a communication and interoperability oriented environment like the Internet, certain technology domains (like web content, as opposed to the Internet Protocol Suite), which have weaker network externalities, may still tolerate proprietary standards.

To devise a standards strategy, a firm will therefore have to take the following into account:

- What is the firm's market power, in relation to its competitors' ?
- How strong are the network externalities for users ?
- What is the value of one's own standard (promote specific technology, ...)

The case studies on web content show that network externalities will increase as a market matures and settles on a standard, thereby making it harder to choose the proprietary way. It was possible for Internet Explorer and Flash to impose themselves as de facto standards in a young market for web content; if they are to be replaced however, that can only happen through an open standardisation process. In the current situation, it is the authors' opinion that new proprietary standards would not be able to supplant the existing ones; indeed the present alternatives, both in the case of IE and Flash, are based on open standards for web content: other browsers adhere much more closely to HTML standards than IE, and the most likely alternative to Flash for video is part of the HTML5 standard.

Bibliography

[1] G. M. P. Swann: *The Economics of Standardization: Final Report*, Department of Trand and Industry. Manchester Business School, December 2000, p. iv.

[2] Interoperable Delivery of European eGovernment Services to public Administrations, Businesses and Citizens: *European Interoperability Framework for pan-European eGovernment Services*; Reference document, Luxembourg, 2004, p. 9, `http://ec.europa.eu/idabc/en/document/3761`.

[3] International Organization for Standardization (ISO): *Graphic technology – Prepress digital data exchange – Use of PDF – Part 1: Complete exchange using CMYK data (PDF/X-1 and PDF/X-1a)*, Standard document, 2001, `http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=29061`.

[4] Internet Engineering Task Force (IETF): *The Tao of IETF: A Novice's Guide to the Internet Engineering Task Force*, Internet-Draft, November 2009, `http://www.ietf.org/tao.html`.

[5] S. Bradner: *The Internet Standards Process - Revision 3*, BCP 9, RFC 2026, October 1996.

[6] G. M. P. Swann: *The Economics of Standardization: Final Report*, Department of Trand and Industry. Manchester Business School, December 2000, p. v.

[7] S.M. Besen and J.Farrel: *Choosing How to Compete: Strategies and Tactics in Standardization*, Journal of Economic Perspectives, American Economic Association, vol. 8(2), pages 117-31, Spring 1994, `http://ideas.repec.org/a/aea/jecper/v8y1994i2p117-31.html`

[8] J. Rohlfs: *A Theory of Interdependent Demand for a Communications Service*, The Bell Journal of Economics and Management Science, Vol. 5, No. 1, Spring 1974, pp. 16- 37, The RAND Corporation, `http://www.jstor.org/stable/3003090`

[9] J. Farrell and G. Saloner: *Standardization, Compatibility, and Innovation*, The RAND Journal of Economics, Vol. 16, No. 1, Spring 1985, pp. 70-83, Blackwell Publishing, `http://www.jstor.org/stable/2555589`

[10] W. Mattli and T. Büthe: *Setting International Standards: Technological Rationality or Primacy of Power?*, World Politics, Vol. 56, No. 1, October 2003, pp. 1-42, Cambridge University Press, `http://www.jstor.org/stable/25054244`

[11] P. Buxmann, T. Weitzel, F. v. Westarp, W. König: *The Standardization Problem - An Economic Analysis of Standards in Information Systems*, Proceedings of the 1st IEEE Conference on Standardization and Innovation in Information Technology SIIT, 1999, http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.37.1056

[12] A. Ventura, G. Perrone, U. La Commare: *Technology standard diffusion and negative network externalities : a lesson from the third Browser War* in the book *Towards new challenges for innovative management practices*, Vol.2, nÂ°1, pp. 45-52, Editors: J. Legardeur, J. Pinho de Sousa, ERIMA Publication, 2008.

[13] P. Windrum: *Back from the brink: Microsoft and the strategic use of standards in the Browser Wars*, MERIT, University of Maastricht, The Netherlands, 2000.

[14] Comparison of layout engines (Non-standard HTML); http://en.wikipedia.org/wiki/Comparison_of_layout_engines_\%28Non-standard_HTML\%29, March 2010.

[15] V. Friedman: *Praxisbuch Web 2.0*, Bonn: Galileo Computing, 2007, p. 541

[16] Acid3 Browser Test; http://www.webstandards.org/action/acid3/, March 2010.

[17] Acid3; http://en.wikipedia.org/wiki/Acid3, March 2010.

[18] File:Layout engine usage share; http://en.wikipedia.org/wiki/File:Layout_engine_usage_share.svg, March 2010.

[19] Frames; http://www.w3.org/TR/REC-html40/present/frames.html, March 2010.

[20] ActiveX; http://en.wikipedia.org/wiki/ActiveX, March 2010.

[21] Web browser market shares in Europe for 2009 Q4; http://www.adtech.com/data/15/browser_EU2009q4_en_middle.jpg, March 2010.

[22] Adobe Systems Inc.: *Adobe Flash Player*; http://www.adobe.com/products/flashplayer/?promoid=DINSO, April 2010.

[23] Wikipedia: *Comparison of layout engines (HTML5)*; http://en.wikipedia.org/wiki/Comparison_of_layout_engines_%28HTML5%29, April 2010.

[24] Verlag Heinz Heise GmbH & Co. KG: *Google schliesst Uebernahme des Codec-Spezialisten On2 Technologies ab*; http://www.heise.de/newsticker/meldung/Google-schliesst-Uebernahme-des-Codec-Spezialisten-On2-Technologies-ab-935165.html, April 2010.

[25] Heise Zeitschriften Verlag GmbH & Co. KG: *Abgeblitzt - Adobe, Apple, HTML5 und die Zukunft von Flash*; c't Article, Hannover, 2010, Nr. 6, March 2010, pp. 194-197.

[26] Heise Zeitschriften Verlag GmbH & Co. KG: *Microsoft mischt auf - Frisch von der Mix-Konferenz: Internet Explorer 9, Silverlight 4 und Windows Phone*; c't Article, Hannover, 2010, Nr. 8, March 2010, pp. 18-19.

[27] Google Inc.: *HTML5-Videoplayer von YouTube*; http://www.youtube.com/html5, April 2010.

[28] D. D. Clark, J. Wroclawski, K. R. Sollins, R. Braden: *Tussle in Cyberspace: Defining Tomorrow's Internet*, IEEE/ACM Transactions on Networking, Vol. 13, No. 3, pp. 462-475, June 2005.

Chapter 3

Economic Aspects of Open Source Software

Fabian Würtz, Marco Creola, Roman Studer

Open Source Software is becoming more and more important. In some area it has established as a serious competitor to proprietary software. This report describes today's economic aspects of Open Source Software. In the first section a brief historical retrospection is given supplemented by some explanations about the different Open Source licenses. Then the advantages of Open Source Software comparing to proprietary software is described. The centre section presents business models in the Open Source field and some real world examples. This section basically shows how companies gain market shares by using Open Source Software these days and how a new market has been generated offering complementary products and services. Furthermore this section explains why it can be economical feasible to invest in the development of Open Source Software even there is no chance of earning money by the software directly. The end of the report discusses some future trends in the Open Source field.

Contents

3.1 Introduction and Problem Statement

Everyone who uses software has once heard about or even used Open Source. But the interpretation of what Open Source stands for may differ. Some people understand Open Source as software where they do not have to pay for. Otherwise the term Open Source is used for every piece of software where the source code has released along with the binaries. But this two properties of Open Source Software (OSS) are not the only ones. To develop, release or use OSS does mean a lot more than it is just for free and you have to provide the source code as well. In a first part this paper will show how OSS evolved within IT history and what it is the common meaning of it today.

With the increased usage of OSS developer of proprietary software began to questioning the whole idea of releasing as free software. In 1976 Bill Gates wrote a letter to all hobbyists in which he complained about software developers which are giving their software for free. Gates main question was: "Who can afford to do professional work for nothing?" [9].

Obviously there are people developing software for nothing. The Internet provides a huge sortiment of software that a user can download without any charges. But why should someone put hard work into a piece of software and then give it away for free? Especially why should a profit oriented company release their software as Open Source? How could a company earn money by developing OSS?

The main section of this paper provides an approach how companies could successfully integrate the development of OSS into a profit oriented business model.

3.2 Definitions and Origin

The first part of this paper shows how the expressions Free Software and Open Source evolved during the last few centuries, the main definitions regarding Open Source and an explanation of the most important Open Source licenses.

3.2.1 Short history of software development

At the time when first software projects have been developed which took place approximately in the sixties the sharing of source code was commonplace. As software was mostly developed by academics and corporate researchers they worked in collaboration and exchanged the source code for free or a nominal charge. In the seventies the development of the first free operating system Unix started.[10]

At the end of the seventies and beginning of the eighties some companies began to distribute their software in form of only machine readable binaries instead of source code. By using licenses the redistribution of software was restricted as well. As a reaction on this trend Richard Stallman created the GNU General Public License (GPL) which should give the developers more rights on their software. As AT&T claimed their rights on the Unix operating system Stallman started a new completely free operating system called GNU Project. To strengthen the basic principles of the GPL Stallman formed the Free Software Foundation.[10]

In the nineties Linus Torvalds created his free operating systems kernel called Linux. With this kernel the so far uncomplete GNU Project was finally a fully functional free operating system. [10]

As some companies tried to use the ideas and principles of free software and adapt it in the commercial software industry they had problems because the paradigms of the free software foundation were to idealistic and radical to completely adapt. To soften this but not loose the advantages of free software the created the expression Open Source and according to their new principles the Open Source Initiative. [10]

3.2.2 Free Software vs. Open Source

The most important groups in the Open Source Community are the Free Software Foundation (FSF) and the Open Source Initiative (OSI). The two following sections are shortly describing this two groups and their direction within software development.

3.2.2.1 Free Software Foundation

The FSF was founded to support the free software movement which is a social movement with the goal to advance software freedom. Based on the ideas of the free software movement and the idealism of Richard Stallman the Free Software Foundation combined the idea of free software with the ethical progression of the current time. The following quotation describes very well how they understand free software: "Free as in Free Speech, not as in Free Beer" [8]. The FSF wants software to be fully available for everyone and completely changeable by everyone.

Free Software Definition

Software is free when the following criteria are within the used license:

- the software can be executed for any purpose
- the software can be studied and changed
- the software can be redistributed
- the software can be improved and newly released

If just one of this items is not guaranteed the FSF defines it as proprietary or non-free.

Copyleft

According to the FSF [8] the goal of copyleft is to guarantee that free software stays free. Every free software has to keep the license it was initially released under.

3.2.2.2 Open Source Initiative

The OSI was founded in 1998 to find a possibility to use free software in a commercial way. The OSI manages the Open Source Definition (OSD), which provides the criteria for open source licences. The OSI also supports the open source community and attends public relations of open source software [28].

Open Source Definition

According to the OSD a open source license must contain the following criteria [29]:

- The software must be redistributable for free, i.e. feeless.
- The license must discriminate neither any persons nor the purpose of use (business or private use).
- The source code must be available as well as the compiled software.
- The source code must be editable and the newly developed software must be distributable under the same license.

3.2.3 Licenses

Software can be categorized by how their source is handled. As figure 3.1 shows software can basically be divided into public domain, open and proprietary software.

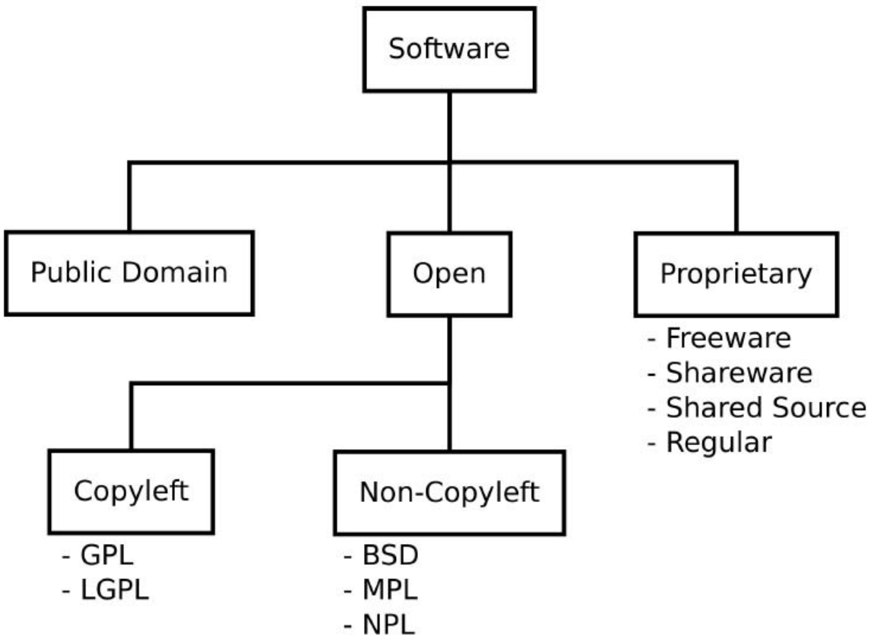

Figure 3.1: Types of Software

Software is under **Public Domain** if the author disclaims all his rights to the software.

Proprietary Software can be divided into freeware, shareware, shared source and the regular proprietary software. Freeware, shareware and regular proprietary software are released without source code otherwise known as closed source. Within this three types they differs by how much they cost and how they are distributed to the users. Freeware as the name stands for it self is free to use. Shareware does have some restrictions like thirty days trial period or some functionalities are only in the full version available. With regular proprietary software there is no version available for free you have to buy the software.

Open Software can be classified by their implementation of the copyleft principle.

The following section gives an explanation of the most used Open Source licenses.

3.2.3.1 GPL – GNU General Public License

The GNU General Public License (GPL) was originally created for the GNU collaboration project at MIT. As a part of the free software movement software released under the terms of the GPL is free software. As it is the first copyleft license it ensures that the software will stay free software. The copyleft license as a kind of mirrored copyright gives everyone the permission to adapt, reproduce or distribute the software as long as it is released under the same license as the original [34].

3.2.3.2 LGPL – GNU Lesser General Public License

The GNU Lesser General Public License (LGPL) is license of the Free Software Foundation. The main difference to the GNU General Public License (GPL) is that if a program uses a LGPL licensed software – for instance as a library – the program can have its own license and must not be distributed under LGPL. This difference was made so that GPL licensed software (e.g. Linux) can interact with proprietary software through a LGPL licensed layer.

3.2.3.3 BSD – Berkeley Software Distribution

The Berkeley Software Distribution is a group of Open Source licenses which have been created at the university of California. Compared to other Open Source licences the BSD gives the users more freedom. For example software under BSD does not have to be made public and is so also deliverable in binary form. This allows better commercial use of the software. The most famous software under BSD is most likely the BSD Operation System and open secure shell application OpenSSH.

3.3 Advantages of Open Source Software

This section describes the main advantages of the use of OSS.

3.3.1 Customisation and its impacts

The most obvious advantage of OSS is made by the definition of Open Source itself. Due the definition the source code of OSS is open. So users can customize the software according to their wishes. Proprietary software is generally closed source and therefore not customisable.

Proprietary software vendors cannot anticipate all customer demands [24]. Therefore they cannot include every possible feature in the product, otherwise costs would explode and the product will become too complex for the vendor [31]. So the ability of OSS to let others adapt the software to their needs is a big advantage for the users.

Thus, a party which adds a feature is interested that the modification is useful for others as well so the code will be included in the main code which is shared by the community [31]. Thus, maintenance costs will be as low as possible because the compatibility to the main code is guaranteed and the piece of code will be tested, bug-fixed and maybe even enhanced which would also be a positive feedback for the own use [31].

3.3.2 Cost and risk sharing

OSS projects encourage the community and its developers to participate in different ways [2]. Mature OSS projects with a large community have the advantage that "the cost and risk of developing the product is distributed among these developers, and any combination of them can carry on the project if others leave" [2].

As shown in the last section developers also share costs and risks not only in the development of OSS but also in maintaining, testing, bug-fixing and extending. They automatically push the overall evolution of the software.

3.3.3 Strict peer review

Code inspection or peer review is obviously very important in software development. Johnson [22] shows that the Open Source community is better in this field than proprietary software vendors. He argues that developers of the community do not care as much about career incentives as developers of a proprietary software vendor (e.g. report code errors of fellow employees).

Lerner and Tirole also mention that in large communities more eyeballs reduce the risk that errors are not detected [24].

3.3.4 Bug fixing and reliability

Kuan [23] compares three OSS with equal proprietary software in regard to program errors, how they are reported and fixed. Her analysis shows that bugs are fixed significantly faster in Open Source projects than in a proprietary environment.

Another study from Crowston and Scozzi [5] analyzes the process of bug fixing. It shows that process sequences are quite short in OSS development. Also the coordination of bug fixing is very spontaneous: developers "who have the competencies autonomously decide to fix the bug" [5].

Thus, bug-fixing is both faster and more effective in Open Source projects than in equivalent proprietary projects.

Bosio et al. [3] study the reliability of OSS. Because failures are faster fixed the community is willing to report more of them and so the more will be fixed. So they assume that this will positively affect the reliability. But the question if OSS is more reliable than proprietary software cannot be answered definitively, i.e. it is too specific.

3.3.5 Development speed

Scacchi [33] also shows that the development speed of OSS is at a faster rate than equivalent proprietary products. He argues the same way as others do: the large community, the independence of developer groups, the fast reaction of the developers et cetera are responsible for high development speed.

3.3.6 Security

One could think that if the source code is available for everyone it must be more vulnerable. Hoepman and Jacobs [19] agree that OSS will be more exposed in the beginning of its development than proprietary software. But "because the source is open, all interested parties can assess the exposure of a system" [19]. Therefore the developers themselves can find bugs and fix them.

As in the last section discussed bug fixing is quite faster and more efficient in Open Source projects than in proprietary projects. Thus, the faster bugs are fixed the faster security patches are available for the whole community the less vulnerable is the software at all [19].

The more mature an OSS project becomes the more insecure code are detected and fixed. Thus, OSS increases its security in the long term so the vulnerabilities in the beginning of the specific development (early days) can be ignored [19].

3.3.7 Innovations

Today OSS is an important source for innovations. Ebert [6] divides them in several dimensions. Innovations can be seen in:

- Process: due to the worldwide collaboration and the incremental development

- Technology: fast adaptation of new technologies
- Quality: fast and effective bug fixing
- Architecture: robustness and modularization
- Standards: HTTP, SMTP, MIME etc.
- Business models: new concepts for distribution
- Marketing: new concepts

According to Ebert [6] “Open source stimulates innovation and ensures that ideas propagate fast to new products.”

3.4 Business Models

Several IT companies (e.g. IBM [21], Hewlett Packard [20]) had already invested billions of US-dollars in the development of OSS in the past years. They do not only invest these amounts but contribute and participate in OSS.

This chapter describes why it could be very profitable for commercial companies to invest and to participate in the development of OSS.

There exist mainly four business models regarding OSS. Each of them is described in a theoretical manner supplemented with at least one practical example.

3.4.1 Integrate OSS in own products/services

This subsection studies the business model of solution providers and software integrators.

3.4.1.1 Economic aspects

According to Riehle [32] a customer often need a complete solution to a given problem regarding to information technology. According to requirements such solutions include hardware, software and services. The services contain all process respective to installation, configuration and at best integration of the components into an existing environment.

Solution providers or system integrators provide such complete solutions. The advantage for the customers is that they have to deal with only one company although several components are involved in the solution [32].

Figure 3.2a shows the described components: hardware, software and services. It also shows the customer demand curve which implicates how many customers would buy the complete solutions at a given price. Typically more customers are interested in a product as cheaper as it is.

If solution providers neither produce the software nor the hardware but rather buy them from a third party they only earn money from selling their services. Thus, solution providers are interested to buy the externally procured products as cheap as possible [32].

Typically OSS is much cheaper than proprietary software [32]. If the OSS does not need any further modifications to be adequate to the customer wishes the usage of OSS can be free of charge at the best.

Figure 3.2b shows how the portions of the components change if OSS is applied to the complete solution instead of proprietary software: because OSS is cheaper than proprietary software solution providers earn more money if they sell the same solution at the same price.

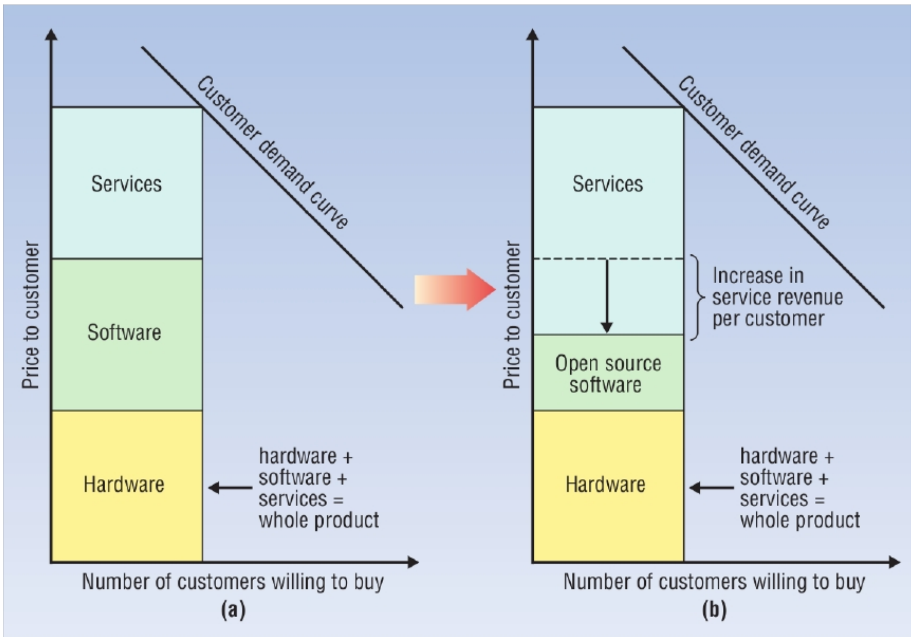

Figure 3.2: IT solutions demand curve [32]

Due to reduction of costs of the software there is another advantage for solution providers. Figure 3.3a illustrates that the service of solution providers causes costs for themselves – logically speaking. As described above solution providers only earn money from the service component less the costs of the service itself. To avoid financial losses solution providers can sell their product at a minimum price of hardware, software and service costs. As Figure 3.3a shows that solution providers will reach a maximum number of customers with this minimum price.

Figure 3.3b shows that solution providers have advantages if they use OSS rather than proprietary software: they can sell their product at a lower price. On one hand they will reach more customers with the minimum possible price. On the other hand they will be at a good position if a price fight among competitors will start because they could sell their product at a cheaper price than their competitors and they would still make profits.

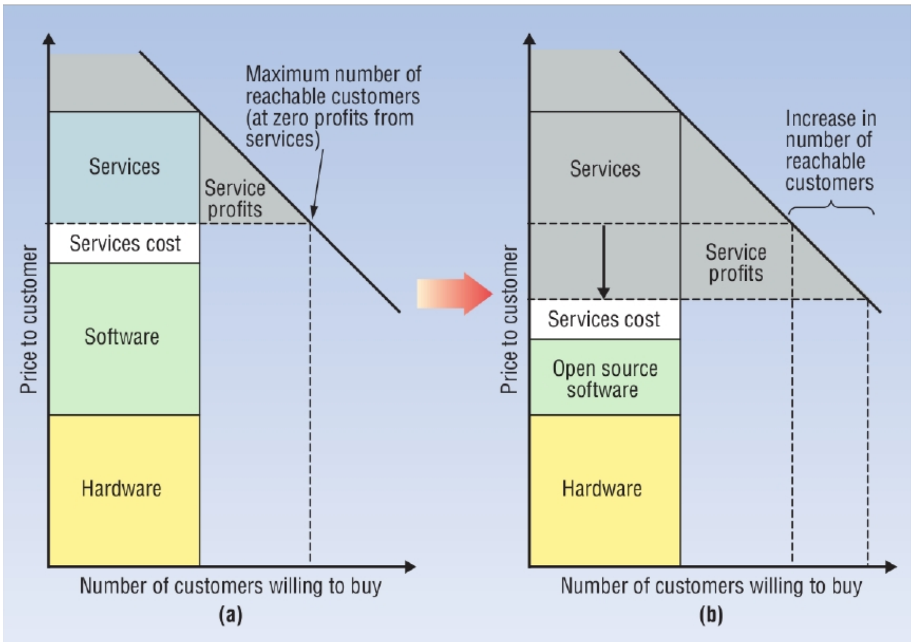

Figure 3.3: Sales margins and numbers of customers [32]

These scenarios show how solution providers could benefit of integration of OSS in their products and services.

3.4.1.2 Example

Meanwhile, there are a number of companies that offer the integration of several open source products and services. The basic idea is to put the different open source components together and offer them as one product from one vendor. This concept is used both at small local companies and big international companies.

Due to it's popularity the American IT company IBM will use as an example for this business model.

IBM has a long experience with UNIX. The company published its own UNIX-based proprietary operating system with the name AIX in 1986. This operating system is available as a bundle together with the required hardware.

In 1999 the company launched its first Linux offers. IBM began actively to participate in Linux development and quickly became the third largest contributor. Meanwhile the investments in the kernel development ascended to one billion dollars per year.

The concept remained the same as the one of AIX. IBM's aim to offer a complete product range that covers everything from hardware to software to support. The richness of the portfolio is shown through the fact that IBM offers more than 500 Linux applications. The programs are a mix of self programmed software and integrated OSS from other vendors.

The advantage for IBM in this business model is that the company is not required to deal with all programming effort itself. Due to the collaboration of many companies,

Table 3.1: Top-three companies working toward the improvement of the Linux kernel [25]

Rank	Company	Changes [absolute]	Changes [%]
1	Red Hat	9351	11.2
2	Novell	7385	8.9
3	IBM	6952	8.3
...	...	...	...
13	HP	765	0.9

universities and volunteers of the community Linux became one of the most popular operating system in the server area.

A disadvantage of this model is that competitors can use the same software as IBM and build their own products, which are compatible with the ones of IBM. One example is the company HP, which has a similar Linux portfolio and a bigger total revenue in the Linux sector, in spite of it invests less into the kernel development than IBM.

Despite other companies benefit through IBM, the half billion dollars net income per year makes the Linux sector a big success for IBM.[13]

3.4.2 Offer complementary products/services

3.4.2.1 Economic aspects

The most basic way for companies to benefit from OSS is to sell additional products or services. Common products are hardware, like mobile devices that work for example with Linux. They are many possibilities in the service area. Many are related with support, in such a way companies offer phone support or training for their open source software. Later this approach will be illustrated with the example of Red Hat.

Companies with open source web applications often host their own software for a fee.

The reason why companies sell complementary products or services instead they sell their software are different. An essential one is the help of the community and other companies. Due to support companies receive a set of advantages like there described in previous chapters. This advantages like the additional patches or the better testing save the companies money and improve their software.

The disadvantage of this model is that companies do not benefit by their software itself. That means that the companies will lose customers who do not need complementary products or support. To avoid this companies often combine this model with dual licensing that will depicted in the next chapter. Another disadvantage is that the success of company depends only on the additional product and services. That means that a company can fail even it has a good product. Furthermore other companies can sell additional products or services without participating at the programming.

Sometimes companies use this business model as loss leader. Due to this strategy the companies hope to enter to an already engaged market. A detailed description of this approach can found in the chapter about the dual licensing business model.

3.4.2.2 Example

A prime example for the provision of additional services to its OSS is the American company Red Hat. Established in 1993, is one of the pioneers in the development of Linux. Meanwhile, as pointed in the table below, Red Hat is the biggest contributor in the Linux development.

Table 3.2: Top-Three companies working toward the improvement of the Linux kernel [25]

Company	Changes [absolute]	Changes [%]
Red Hat	9351	11.2
Novell	7385	8.9
IBM	6952	8.3

Despite of Red Hat's different activities, as pointed in the pie chart below (Figure 3.4), the most important product is the Linux distribution "Red Hat Enterprise Linux". Since the company is not eligible to sell Linux as closed operating system it sells so-called subscriptions instead. This subscription includes the downloading of the operating system and the supply of packaged security updates. Depending on the price the subscription contains different ways of support. The range is comprised of simple web support to 24 hours phone support.

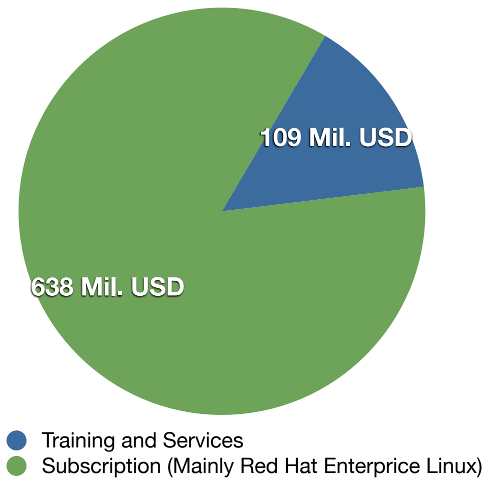

Figure 3.4: Red Hat: Revenue by sector. [26]

The foundation of "Red Hat Enterprise Linux" is Fedora. Fedora is a Linux community operating system maintained by Red Hat and volunteers. It allows Red Hat to test new software and later integrate it into the enterprise version. The Advantage for Red Hat is the free help from the voluntary community members. This help provides packaging and debugging. Of course, the benefit is mutual: While the community receives a complete free and updated operating system, Red Hat gets a stable and well tested enterprise distribution. One can therefore speak of a win-win situation. In addition it also increases the degree of popularity of Red Hat and its products.

Nevertheless, this business model brings some disadvantages. It is possible to offer replicas of "Red Hat Enterprise Linux" such as CentOS. The only difference is the lack of support. So Red Hat does not benefit from customers, who do not need support. In addition other companies offer subscriptions for the replicas.

Another challenge for Red Hat is the interaction with the community. It would have to influence the community without directions in order to keep its volunteers. The company has solved this problem with a compromise. In the main Fedora council are 4 employees of Red Hat and 5 elected members of the community. [7]

Figure 3.5 shows that Red Hat is a success story. Even if the total revenue is small compared to Microsoft ($58,437 billion/2009)[27] or Apple ($42,905 billion/2009)[1], it proves that the business model can be practised very successfully.

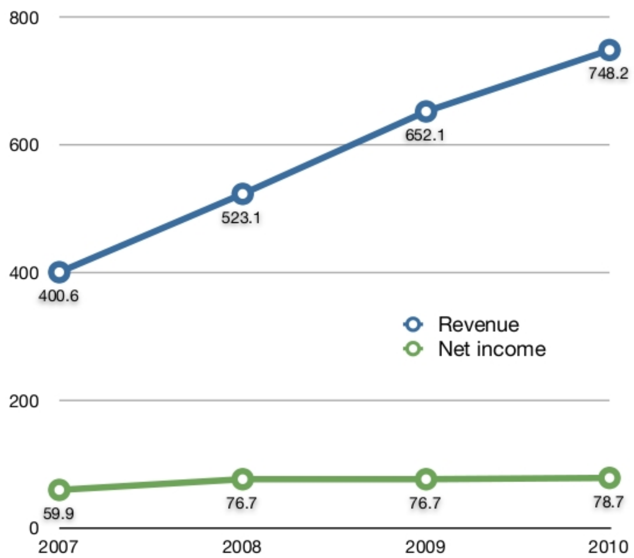

Figure 3.5: Red Hat: Revenue and net income.[18]

3.4.3 Release own code

This subsection shows how software vendors can profit of releasing their code to the Open Source community.

3.4.3.1 Economic aspects

If a manufacturer of proprietary software only plays a marginal role in a specific software market it could be economical feasible to open the own software for the community or at least some parts of it. With a dual licensing software model vendors can off their software in two editions: the first edition is open source and the second edition is proprietary. One could either release the proprietary version with more functions or with more rights to the customer, e.g. customers can use the edition in commercial projects (see the example in the next section). [24][32]

The dual licensing model makes also sense for software vendors, which want to enter a new software market. If the entry barrier of a market is very high (e.g. the market has one big leader) this model can be the most profitable and the only one to succeed. [32] Because the established market leader dictates the price and the costs of the development of competitive software are very high, one must share the risks and costs of the development what is achievable by the development of OSS. [32]

Also if software vendors want to make their software interoperable with a maximum number of other software and hardware components, e.g. drivers and interfaces, releasing parts of the code to the Open Source community will lower the costs to a minimum to achieve this aim. [24]

Other advantages of releasing the code are directly taken from the overall advantages of OSS as described in a previous chapter: the Open Source community transforms the software into a better product by better bug-fixing, by extensive testing [23] and by strict peer review [22].

3.4.3.2 Example

A prime example of the dual licensing business model is MySQL. The database was originally developed by the Swedish company "MySQL AB" that was founded in 1995. The software was originally developed as a replica of mSQL and quickly developed into one of the leading open source databases. To benefit from the database the company has released the software under the GPL and a property licence.

The most important advantage of the the property version is that this version can be used in commercial projects. An other advantage is that changes must not be re-published.

The business model has made the rapid distribution of the software possible. Meanwhile MySQL is one of most popular databases. Due to the open source version the company received free bug reports and ideas for improvement from the community.

But there are also disadvantages of this model. Due to copyright issues it is not possible to integrate patches from community members into the property version. To keep the versions identical the developers do not accept patches with copyright from volunteers at all.

Even if the company has never released exact sales figures or profit reports, it is likely that the company was successful with their product and the dual licensing business model. Until 2008 the company enhanced to about 400 employees in 25 countries to one of the largest open source companies. Also the fact that the company Sun paid 1 Billion dollars to take over "MySQL AB" emphasis the success. Even through Oracle, a company that sells property databases, bought Sun in 2010, the dual licensing business model was continued.

3.4.4 Use OSS in-house

The following subsection describes why companies should participate in the development of OSS if they use the mentioned software for internal processes.

3.4.4.1 Economic aspects

Before companies build software from scratch for internal use they should check if existing and mature OSS fits their wishes. The missing features can be added with lower costs than building the whole software. Because mature OSS is developed by several developer groups and as described in a previous section the costs of OSS development are shared within these groups. [2]

The possibility of adding features to the software implicates the next economical aspect: OSS can be adapted to own needs, i.e. OSS is completely customisable.

OSS development also minimise the risks of the development itself. Similar to the costs also the risks are shared within the developer groups. [2]

A further advantage of the usage of OSS for internal processes is that the company is more independent from an external third party. [4]

3.4.4.2 Example

One of the well-known example of in-house usage is Eclipse. This Java development environment was developed in the Canadian studios of the IT company IBM. The software was created to as a powerful development tool for their own Java projects. Therefore it was not a customer order but rather a co-product the in-house usage.

Due to in-house usage, there was no absolute need to earn money with this product. Nevertheless the advancement of Eclipse was still important for the company, inasmuch the program proves to be useful for further projects, too.

In order to attract more developers and other companies to join the development of the program, IBM decided on 7th November 2001 to publish the software under a free license. As license the company was choosing their own EPL (Eclipse Public License) that allows in comparison to the GPL the distributing of commercial extensions. The EPL is not compatible with the GPL. The concept of open licensing proved to be successful. Today, a large number of well-known companies, such as Oracle, Nokia and SAP, are part of the programming. To advocate this development IBM founded in 2004 the Eclipse Foundation that coordinates the development.

To sum up, IBM has managed to develop a powerful development environment for their own internal use, without having had to pay the sole cost of the development. The only drawback of this model is that now also competitors of IBM can use this tool for free.

3.5 Trends

Market studies show that OSS has become an important issue to companies as proven in the application area and in development section. In 2008 the market research company Gartner researched in 274 companies of different sizes in North America, Europe and the Pacific area. The result was that 85 per cent of the considered companies use OSS. The other 15 per cent plan the usage for the future.[11] Research company IDC gets similar results in it is study of 2006, 71 per cent of 116 companies use OSS.[12]

But what does that increasing usage of OSS mean for its contributors? It is likely to expect profits for the companies due to this rising demand. In addition it is likely that more companies will try to jump on the bandwagon and release their software as open source as well. Might raise the upcoming demand for OSS due to public open source initiatives of some countries.

Another interesting development is that some open source companies like Red Hat[14], Novell[15], Alfresco[16] or Windriver[17] benefit from economic crisis. In the last months, they enlarged the customers figures or business volume. Rather this depicts a general trend or simply current individual cases, only the future will show.

Even if the four presented business models contains at moment the most of the products of open source companies, a constant modification of the product is likely. Especially in the area of the additional products and services. In spite of completely new business models are at present indeterminable.

Even the rise of the OSS it does not look that proprietary software will disappear. Examples like the "app stores" from Apple and Google show that the simple direct sale business model of proprietary software vendors is still working very successfully.

In the last years the acceptance and usage of open source was rising up. Accordingly this development will continue in the following years it makes sense to speak of a positive trend of open source and it contributors. In spite of a domination of OSS and the end of the property software is unlikely. More likely both business models will exist side by side in the future.

3.6 Summary

OSS has in recent years always gain distribution. Meanwhile more and more companies join open source projects or release their own software as open source. But what is the motivation and how can companies benefit from OSS?

The paper has depicted that there are different reasons to program OSS. A main advantage for the companies is the collaboration with the community. The cooperation can save money and has less risks. In addition the community can improve the bug fixing and the security of a program. Aside this there are further reasons.

Meanwhile companies have found different ways to benefit from OSS. The paper has divided this business models into four categories. In the first category companies integrate OSS in their own products or services. In the second one are companies that offer complementary products or services to their OSS. The third category are companies that earn money by selling their code with dual licensing. The last category are companies that use a in-house business model. To depict the models the paper has inspect a successfully example for each category.

As trend the authors of the paper believes that OSS will continue its rise. It is likely that public open source initiatives will boost this development. In spite of it is unlikely that proprietary software will disappear. More likely both business models will exist side by side in the future.

Even if this short paper has already depicted an overview about the economic aspects of open source there are more interesting topics. For example a continuative paper could compare successful an unsuccessful open source companies to find out the formula for success in the open source business area.

Bibliography

[1] Apple Inc.. *Revised Request Quarterly Income Statement*, available at http://phx.corporate-ir.net/External.File?item=UGFyZW50SUQ9Mjc1MjN8Q2hpbGRJRD0tMXxUeXBlPTM=&t=1, January 2010.

[2] Bessen, James. *Open Source Software: Free Provision of Complex Public Goods*, available at http://www.researchoninnovation.org/opensrc.pdf, July 2005.

[3] Bosio, Diana; Littlewood, Bev; Newby, Martin; Strigini, Lorenzo. *Advantages of open source processes for reliability: clarifying the issues*, available at http://www.csr.city.ac.uk/people/lorenzo.strigini/ls.papers/_02_open_source/OSSBosioetal.pdf, February 2002.

[4] Bernd Brügge et al. *Open-Source-Software: Eine Ökonomische und technische Analyse*, Springer Verlag, 2004.

[5] Crowston, Kevin; Scozzi, Barbara. *Coordination practices within floss development teams: The bug fixing process*, available at http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.84.2366&rep=rep1&type=pdf

[6] Ebert, Christof. *Open Source Drives Innovation.* In: IEEE Software, pp. 105-109, May/June 2007.

[7] Fedora Project. *Board*, available at http://fedoraproject.org/wiki/Board

[8] Free Software Foundation. Available at http://www.fsf.org/about/

[9] Gates, William Henry. *An Open Letter To Hobbyists*, 1976, available at http://www.blinkenlights.com/classiccmp/gateswhine.html

[10] Gutsche, Jörg. *Ökonomische Analyse offener Software*, 2005.

[11] h-online.com. *Gartner: Open Source is pervasive*, available at http://www./newsticker/news/item/Gartner-Open-Source-is-pervasive-738189.html

[12] heise.de Open Source. *Studie: Open Source gewinnt an Fahrt*, available at http://www.heise.de/open/meldung/Studie-Open-Source-gewinnt-an-Fahrt-152679.html

[13] heise.de Open Source. *HP und IBM scheffeln Geld mit Linux*, available at http://www.heise.de/newsticker/meldung/HP-und-IBM-scheffeln-Geld-mit-Linux-73437.html

[14] heise.de Open Source. *Red Hat: Wachstum in der Wirtschaftskrise*, available at http://www.heise.de/open/meldung/Red-Hat-Wachstum-in-der-Wirtschaftskrise-209597.html

[15] heise.de Open Source. *Gemischte Gefühle bei Novell*, available at http://www.heise.de/open/meldung/Gemischte-Gefuehle-bei-Novell-201765.html

[16] heise.de Open Source. *Alfresco verdoppelt Jahresumsatz*, available at http://www.heise.de/open/meldung/Alfresco-verdoppelt-Jahresumsatz-210153.html

[17] heise.de Online. *WindRiver steigert Umsatz und Gewinn*, available at http://www.heise.de/newsticker/meldung/WindRiver-steigert-Umsatz-und-Gewinn-205133.html

[18] heise.de Open Source. *Red Hat auf Wachstumskurs*, available at http://www.heise.de/newsticker/meldung/Red-Hat-auf-Wachstumskurs-963696.html

[19] Hoepman, Jaap-Henk; Jacobs, Bart. *Increased Security Through Open Source.* In: Communications of the ACM, Vol. 50, No. 1, pp. 79-83, January 2007.

[20] HP. *Open Source and Linux from HP*, available at http://h20338.www2.hp.com/enterprise/cache/599999-0-0-0-121.html

[21] IBM developerWorks. *New to Open source*, available at http://www.ibm.com/developerworks/opensource/newto/

[22] [4] Johnson, Justin P. *Collaboration, peer review and open source software.* In: Information Economics and Policy 18, pp. 477-497, 2006.

[23] Kuan, Jennifer. *Open Source Software as Consumer Integration into Production*, available at http://pascal.case.unibz.it/retrieve/3300/Jenny+Kuan+-+Open+Source+Software+As+Integration+into+Production.pdf, October, 2000.

[24] Lerner, Josh; Tirole, Jean. *The Economics of Technology Sharing: Open Source and Beyond.* In: Journal of Economic Perspectives, Vol. 19, No. 2, pp. 99-120, 2005.

[25] Linux Foundation. *Linux Kernel Development*, available at http://www.linuxfoundation.org/publications/linuxkerneldevelopment.php, April 2008.

[26] Linux Magazine. *Red Hat legt kraeftig zu*, available at http://www.linux-magazin.de/NEWS/Quartalsbericht-Red-Hat-legt-kraeftig-zu

[27] Microsoft. *MSFT Annual Report 2009:*, available at http://www.microsoft.com/msft/reports/ar09/10k_fh_fin.html, October 2009.

[28] Open Source Initiative. *About Open Source Initiative*, available at http://www.opensource.org/about

[29] Open Source Initiative. *Open Source Definition*, available at http://www.opensource.org/docs/osd

[30] Open Source Initiative. *Open Source Licenses*, available at http://www.opensource.org/licenses/alphabetical

[31] Perens, Bruce. *The Emerging Economic Paradigm of Open Source*, 2005, available at `http://perens.com/Articles/Economic.html`.

[32] Riehle, Dirk. *The Economic Motivation of Open Source Software: Stakeholder Perspectives.* In: Computing Practices, pp. 25-32, April 2007.

[33] Scacchi, Walt. *When Is Free/Open Source Software Development Faster, Better, and Cheaper than Software Engineering?*, available at `http://www.ics.uci.edu/~wscacchi/Papers/New/Scacchi-BookChapter.pdf`, June 2004.

[34] Smith, Brett. *A Quick Guide to GPLv3*, available at `http://www.gnu.org/licenses/quick-guide-gplv3.html`

Chapter 4

Standards and Models for Expressing and Monitoring SLAs

Patrick Leibundgut, Alexander Filitz, Damian Schärli

Service Level Agreements are used to define the Quality of Service. The SLA defines a contract between a service provider and a service user. There are several different models for expressing and monitoring a service. In this paper the most common ones are introduced. After a short description of the language and an overview of the actual development situation, there is a comparison of all the languages. Some languages are defined more to express the Service Level and only a few provide a support for the monitoring part. For different services you need a different language. A common part of most of the languages is the XML, which allows extending the language and it is very efficient. At the moment there is not one official accepted standard or implementation that would make it possible even for a small business to support SLAs in an automated manner. Large companies are trying to force their language to an international standard. At the moment most of the SLAs are written in natural language. It still needs some development to enlarge the financial and investment advantages of using XML based languages.

Contents

4.1 Introduction

4.1.1 What is an SLA

Nowadays the demand of connection and service with a certain Quality of Service (QoS) rises with the growth of the internet and its users. These services need a contract in which the contractual bases are well-settled in a clear manner that describe parties' responsibilities and duties. The Service Level Agreement (SLA) is an possible solution to solve these problems. An SLA is a contract or a part of a contract made between a service provider and a client that assigns penalties to undesirable behaviour on a given service [14]. The SLA is a product of all the points that the service provider and the client worked out and want to fix in the contract for measuring or monitoring reasons. It is important to know that only mentioned parameters in the SLA are part of the contract, and any party can only argue against these agreed terms.

Some typical parameters that are in an SLA, for example, defining a service level with expressing quantitative parameters, regulation of the fees and penalties in case of not respecting the service agreements and giving the opportunity to measure and monitor predefined levels of service [15]. To get an impression how an report of an monitored process could look like, there is the following picture. On this graphic there is an overview of a performance measure with its acceptance levels. The horizontal line marks the performance agreement of the SLA which is set to 85%. The blue bar indicates acceptable performances and the yellow and orange ones are unacceptable and critical performances respectively.

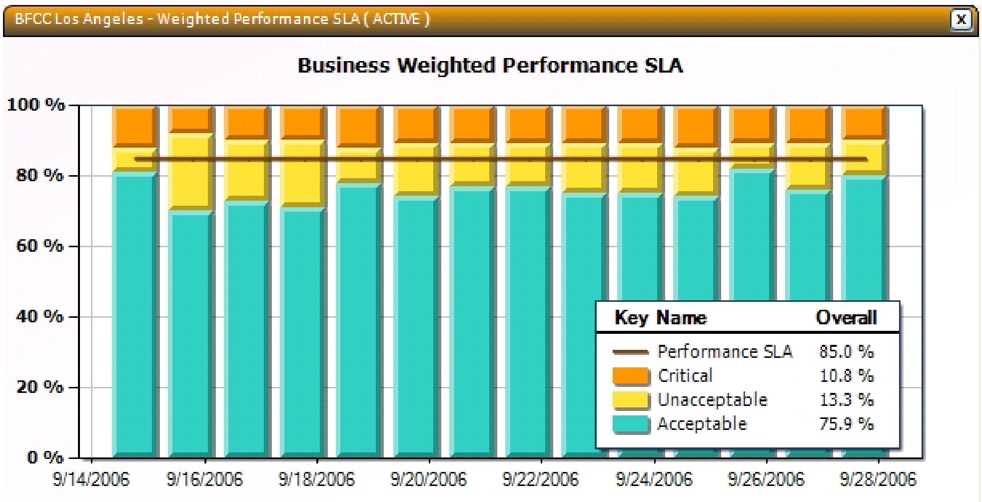

Figure 4.1: Example of an SLA report[1]

[1]Source: `http://www.evidant.com/includes/images/sp_shots/SLA_Perf.png`

4.1.2 Contents of SLA

The following section treats the contents of an SLA. An SLA is not exactly defined in a book. It is very adaptive and it is difficult to define an exact shape of the contract. That is why the following specification has to be understood as a possible SLA's content. In a real service situation, both parties have to sit together and work out every single aspect that they want to have in their contract and want to monitor and they have to set the service level.

In an SLA there are several different parts. In a first section there are the general definitions. This is the duration of the SLA. You also find the involved parties. These are the service provider and the client. You should specify every point in a detailed manner including things like the duration of an SLA, whether there are any breaks or if it is possible to have a service delegation.

In an other part there are the juridical and financial aspects.Exactly in this part of the SLA the consequences of not fulfilling the service levels and quality are defined. It is possible to fine a party or to describe the way to treat contract violation. In the financial aspects is defined how the service is paid. There are several different methods to pay the service. Most commonly it is a combination of flat price, periodic payment and direct payment to the actual service.

One important issue is to write down the service levels. These levels contain *availability* (ratio between the time in which the service is available and the time in which it could be available. For example a runnig server could run 24 hours a day and seven days a week but is offline for three minutes a week to restart.), *service time* (working time), *escalation* (determination of the process in case of an error inclusive time of response. In other words what happens if there is an error and how long does it take to resolve it.), *time of response* (time in which the service provider has to act to solve a problem considering the priority), *documentation* (a description the way that the document was created/processed), *security* (security levels and treatment of personal data) [15]. They are often expressed in an similar way. To achieve all those requirements this is often done with metrics. The following section is about those metrics.

An SLA has typically some common metrics to define key measures. To get an impression of what this could be, here is an example for an IT service management such as call center or service desk [4]:

- Abandonment Rate: This is the percentage of abandoned calls which were in the waiting queue to be answered.
- Average Speed to Answer: This is the average time to answer a call by the call center. It is measured in seconds.
- Time Service Factor: This is the percentage of the answered calls in a predefined time frame. It could be something like 85% in 1 minute.
- First Call Resolution: This is the percentage of the incoming calls that can be solved without the need of a callback of the helpdesk. The problem can be fixed by the call center employee directly.

- Turn Around Time: This is the time to complete a certain task as part of the service.

This is only one possible selection of measures for a call center. For other areas you need other measures. For example for a contract about a computational resource it could be something like time response in a network or the time to make a query in a database.

4.2 Standards and Models to express SLA

4.2.1 Aspects of an SLA

In the following list you find important requirements for an SLA language and SLA specifications [16]:

4.2.1.1 SLA Language Aspects

As mentioned before, SLAs need a concrete representation. In order to express an SLA it is needed to use a given language. Such language can be a natural language, a technical language or just a combination of both. Requirements for the SLA language are as follows:

Expressiveness The language must express all types of SLAs. This implies that all requirements related to an SLA are also important for the language.

Understandability The language has to be structured and easy to understand.

Precision The language has to be precise, the semantics of the language have to be defined and it has to be clear in its meaning.

Restrictiveness The language excludes SLAs out of requirements specifications.

Ease of use The language has to be easy to write/express. Perhaps a tool can support the writing of such language.

Power The language is defined once and can be used for different SLAs.

Automatability The language has to be in a way to be able that a tool can process it and take it as a source for the monitoring of the tool.

Analysability The semantics of the language must be compatible with expressing the true requirements of the client to avoid exploitability.

4.2.1.2 SLA Specification Aspects

The quality of the specification has an effect on the usability of the SLA language. Here are some requirements for the specification:

Completeness The specification has to define an SLA language meeting all requirements.

Understandability The specification must define the language in an understandable way.

Precision The specification must define the language in a precise way.

Automatability The specification should be defined in a way to assist the developement of tools.

4.2.2 WSLA

4.2.2.1 Overview

WSLA is the acronym for Web Service Level Agreement. The XML[2] based document defines agreed guarantees from the provider to the service user. WSLA is not quite clear and exact. It can be abstract to define parameters which are necessarily needed to define the service. This abstraction helps to use the WSLA Agreement for specific domains and technologies. Due to this it is possible to use automated management frameworks which support controlling the SLA contract and inform one involved party in case that there are is a violation of a SLA parameter, for example, if the reaction time of a service must be under 2 milliseconds and the service takes 10 milliseconds [6].

The design goal of the WSLA definition is to provide a flexible and formal language which can interoperate with electronic commerce systems and shows SLA informations about different parties. The measurement of the defined parameters is not part of the core specification, they are defined in the standard extensions [7].

4.2.2.2 Language Description/Structure

The language was developed by a research team of IBM until 2003. The agreement is written with XML and the specifications are very detailed as it is shown in the next paragraph [6].

The picture in Figure 4.2 shows the concept of the structure of an agreement. As described in the overview of the language, WSLA files are written in XML. The three main contents are the parties, the obligations and the service definition. The section of the

[2]Extensible Markup Language: Common standard to express well defined documents in internet based communication.

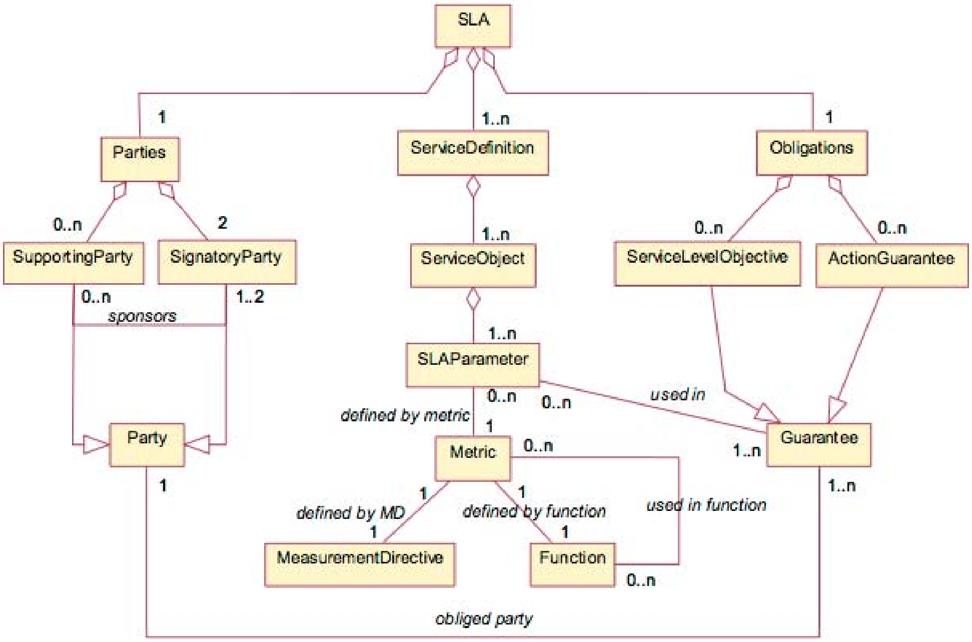

Figure 4.2: Overview of main WSLA concepts [6]

parties describes the offering party, the using party and under certain circumstances the supporting party. The service definition can have more than one service object which is an abstract of service like WSDL/SOAP operation. This service objects themselves have one or more SLA parameters. They are responsible for defining the service under consideration of metrics, which are responsible for the measurement of the service operation. In the part of obligations the guarantee is defined like "the service object takes less than 2 milliseconds". So this is linked to previous mentioned SLA parameters [6].

4.2.2.3 Development Status

The first specifications of WSLA came out in 2003. Since then neither IBM nor another company or group published a newer version. So there is no further development in progress [6].

4.2.2.4 Example Scenarios

For a better understanding of WSLA imagine a simple Web Service on the internet with one operation. The WSLA would now define parameters which are related to this operation. For example the time the service has to reply should be under two milliseconds. Because of the advantage of the XML generated agreement a third party should be able to automatically resolve the agreement for its correctness. The definition also allows the user to look which guarantee he has if he uses the service and which not. Moreover the offering party of the service and the users which are allowed to use the service are defined.

4.2.3 SLAng

SLAng is an acronym for Service Level Agreement Language. A detailed specification can be found on the homepage of UCL[3]. The main idea of SLAng is to define an end-to-end Quality of Service (QoS) [1].

4.2.3.1 Overview

SLAng as implementing QoS standards can be used with several technologies in distributed systems like WSDL/SOAP[4] or systems using CORBA specification [1]. It was first described by Davide Lamanna mainly for use in the ASP(Application Service Provision) domain. The specification is based on the traditional N-tiered application service architecture as seen in figure 4.3. In this domain there are two main SLAs available:

Horizontal SLA Between two services on the same application height, e.g. Network-Network connection. This allows to define a contract between to items on the same height in the application structure but on different application execution places.

Vertical SLA Between two services on the same application stack, mainly in the same scope of the application.

In Figure 4.3 the architectural components of a distributed system are shown as nodes. The edges describe the opportunities of Service Level Agreements between two parties which can be expressed with SLAng. This component model was the background developing of this Service Level Agreement Language. The development of a language which is usable for a user near side like application and for a more technical part like for example, in the Web Services is the main goal of the SLAng language [1]. The SLA is also defined very abstract but with a clear semantic. This supports the monitorability of the Service Level Agreements in an automated manner [16].

4.2.3.2 Language Description

The SLAng synatax uses an XML Schema to define an SLA. This allows to combine the SLA with BPEL and WSDL to a complete e-Business automation solution. BPEL[5] and WSDL[6] defined services are also written in XML Schemas and this allows to interact with the WSDL or BPEL service [1]. The content includes generally the following elements and attributes:

End-point description Contains information about the contractors, location and facilities.

[3]See: http://www.ucl.ac.uk

[4]Defines Web Services over HTTP with XML

[5]Business Process Execution Language: Allows to run Business Processes running on a generic engine without programming itself.

[6]Web Service Definition Language: Describes an available Service and how to use it.

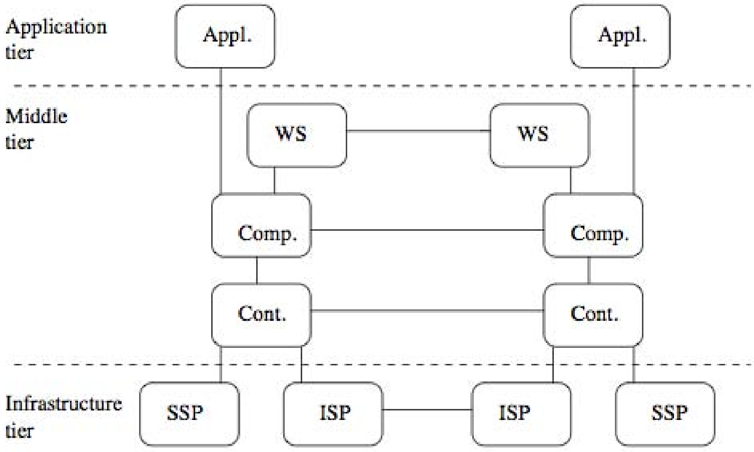

Figure 4.3: Service provision in three-tiered architectures

Contractual statements Contains information about the agreement, charging clauses, start date and so on.

Service Level Specificaiton Are also called SLS. This are techinical defined Quality of Service (QoS) descriptions and metrics.

4.2.3.3 Development Status

The SLAng language allows to monitor the service. But the effectiveness of this monitorability is part of the future work which has to be done. Additionally there is some future work on a toolkit for the service composition in ASP to analyze which SLA are allowed and useful [1].

4.2.4 WSOL

4.2.4.1 Overview

WSOL refers to Web Service Offerings Language which is based on XML syntax and obviously compatible with WSDL. It supports to specify functional behaviour like pre, post and future conditions as well as QoS and authorization rights. Additionally a person can define management terms about *e.g.* penalties or prices. A WSOL offering is similar to an SLA or a simple contract. The developers of WSOL claim that the main benefits of this language are the expressive power, reducing run-time overhead and the focus on management services.

4.2.4.2 Language Description

The main concept of WSOL is to describe classes of service of one web service. So classes of a web service contain the same functionality, but vary *e.g.* in guaranteed response time, penalties etc. You can also define which hardware has to be used to run the service of this class and different classes of service can define their own payment structure. This allows a company to easily provide a single web service to different consumer segments. And this concept also reduces run-time overhead, because the classes can be reused and have less complexity than custom made SLA's [12].

As mentioned before the syntax of WSOL is based on a XML schema and five constructs namely *constraint*, *statement*, *constraint group (CG)*, *constraint group template* and *service offerings*. Constraints are boolean expressions which are evaluated before or after an operation starts or at a specific time. WSOL distinguishes between functional constraints, QoS (non-functional) and access rights. Statements define information about the belonging class of service. This could be *e.g.* price, penalty and external operation calls. Constraint groups are composed by constraints and statements which can be reused for other class definitions. Constraint group templates differ from normal CG's by using abstract parameters. Service offerings are syntactically the same as CG's, but they dont have to be nested. WSOL introduces an accountingParty element for service offerings, which is responsible for monitoring the offering. In the end a consumer buys a service offering and not a CG. With this model it is also possible to have third parties controlling the service offering as you can see in Figure 4.4 [12].

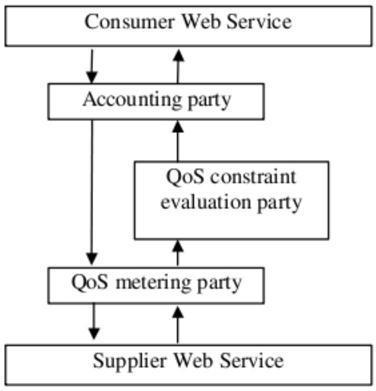

Figure 4.4: WSOL QoS management of third parties

4.2.5 WS-Agreement

4.2.5.1 Overview

The final specification of WS-Agreement is a publication of the Global Grid Forum 2007 in context to other WS-* specifications. In this case WS simply refers to Web Services in general. Normally a service provider can not provide global information about quality, availability and similar parameters, but instead he has to deal with real-time information like actual workload and other environment settings. When an agreement finally is accepted, it is necessary to monitor the compliance to inform the service consumer about a violation of the terms in an agreement. WS-Agreement addresses all these problems and contains mechanisms to express service requirements, service guarantees and service descriptions and it also allows to observe their compliance. The goal is to standardize the overall agreement structures with different types of terms and also a set of operations to create and monitor these agreements [10].

4.2.5.2 Model Description

Basically a WS-Agreement system consists of two layers. An agreement layer and a service layer. The agreement layer is used to create and monitor the agreements and the service layer provides the application specific service. As soon as an agreement is accepted in the agreement layer the underlying service layer is controlled by the rules of the agreement. The binding between these two layers is defined in the agreement itself. An agreement document is written with a XML scheme according to the W3C standard[7], which is better described in the following section [10].

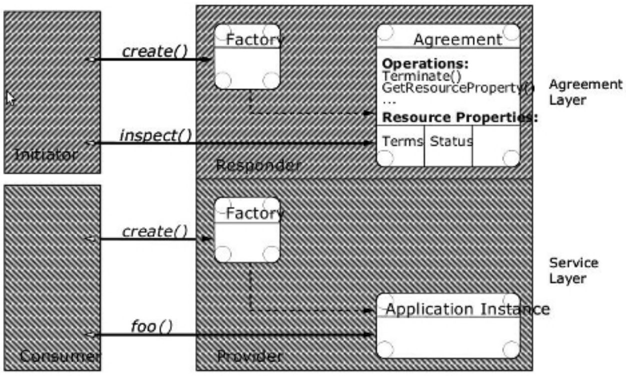

Figure 4.5: WS-Agreement Layered Model

4.2.5.3 Agreement Structure

The structure of an agreement is composed by three main parts as you can see in Figure 4.6. While the name is optional the other two sections are mandatory. In the context section lies the meta-data, like participants and lifetime information. The main part of the agreement is the term section, where you describe the requirements and guarantees to fulfil the contract. Terms are distinguished between Service Description Terms (SDT) and Guarantee Terms. Service description terms define the functionality which will be offered by the agreement. It is possible to define multiple SDT's, because a service could contain different functionalities or different services are used by one agreement. The most important part for our topic are the Guarantee Terms. They assure a certain level of service quality and availability to the service consumer and include Service Level Objectives (SLO). Every guarantee term also includes qualifying conditions and business values. Qualifying conditions express rules on external conditions like access daytime, rate or other terms a service consumer must satisfy. Business values describe the weight or the significance of an SLO to a service consumer, provider. Like with the SDT's one

[7]http://www.w3.org/standards/xml/

can define any number of SLO's for an agreement which can be composed objectives as well [10].

Figure 4.6: WS-Agreement Structure

4.2.5.4 Development status

The specification itself is complete and closed so far. Nowadays the number of different implementations is increasing successive. For example there is the WSAG4J[8] project which implemented a generic WS-Agreement framework software package. Like almost all of the existing implementations WSAG4J is written in Java. A complete list of implementations of the WS-Agreement specification is available at [21].

4.2.5.5 Example Agreement

```
<wsag:Agreement AgreementId="1">
   <wsag:Name>Sample Agreement</wsag:Name>
   <wsag:AgreementContext>... </wsag:AgreementContext>
   <wsag:Terms>
    <wsag:All>
     <wsag:ServiceDescriptionTerm wsag:Name="Sample Agreement"
        wsag:ServiceName="Sample Service">
       ...
     </wsag:ServiceDescriptionTerm>
     <wsag:GuaranteeTerm Name="Sample Agreement" Obligated="
        wsag:ServiceRoleType">
      <wsag:ServiceScope ServiceName="Sample Service">...</
         wsag:ServiceScope>
      <wsag:QualifyingCondition>...</wsag:QualifyingCondition>
      <wsag:ServiceLevelObjective>...</
         wsag:ServiceLevelObjective>
      <wsag:BusinessValueList>...</wsag:BusinessValueList>
     </wsag:GuaranteeTerm>
    </wsag:All>
   </wsag:Terms>
</wsag:Agreement>
```

[8]http://packcs-e0.scai.fraunhofer.de/wsag4j/

4.2.6 WSPL

4.2.6.1 Overview

WSPL refers to Web Service Policy Language and therefore it is meant to be a standard language to express web-service policies. The language was created by Sun Microsystems Laboratories after some use cases and needs came up in a public forum. There are many different parameters and features of a web service that a policy can manage. For example authentication, Quality of Service or service specific options. WSPL supports any policy by using standard data types and functions, merging of two different policies to one intersectional policy and comparison of policy parameters better than single equality match.

4.2.6.2 Language Description

The language is a subset of the XACML [22] standard. In general a policy consists of one or more rules, which are ordered descending with the most preferred first. Each rule describes to which part of the web-service the rule is applied to and includes a set of predicates which all have to be satisfied to make a rule and so a policy valid. As you can see in Figure 4.7 there are three rules for the aspect "Quality of Protection" in a specific web service. The first rule, the most preferred one, states that the signature algorithm must be "RSA-SHA1" and a key length greater or similar to 2048 must be used. In the second rule the satisfying condition for the key length is only 1024 if the source domain is equal to "EXAMPLE.COM". And the last rule says that for the source domain "MY.EXAMPLE.COM" nothing is required [11].

```
Policy (Aspect = "Quality of Protection") {
  Rule {
    Signature-Algorithm = "RSA-SHA1",
    Key-Length >= 2048 }
  Rule {
    Signature-Algorithm = "RSA-SHA1",
    Key-Length >= 1024,
    Source-Domain = "EXAMPLE.COM" }
  Rule {
    Source-Domain = "MY.EXAMPLE.COM" }
}
```

Figure 4.7: WSPL policy rules

All policies for a specific web-service are combined in a policy set. If a web-service needs more than one policy set it is possible to combine them in a second level inside the top policy set. Figure 4.8 shows such an example structure.

A major characteristic of WSPL is the ability to support negotiating of policies between the service consumer and the service, or between two services. This is done by merging the two policies to find out if they match or not. Three steps are necessary to merge two policies. First, the targets of the policy set must match. Second, pairs of all possible rule combinations are calculated. Third, each pair is combined to one rule, if that is not possible, the pair gets deleted. The surviving rules build the merged policy, if all pairs are deleted, the merge fails [11].

```
PolicySet (target=<port type>) {
  PolicySet (target=<operation/message>) {
    Policy (target=<aspect>) {
      Rule {
        <predicate>, ...
      } ...
    } ...
  } ...
  Policy (target=<aspect>) {
    Rule {
      <predicate>, ...
    } ...
  } ...
}
```

Figure 4.8: WSPL policy set

4.2.7 ODRL-S

4.2.7.1 Overview

ODRL means Open Digital Rights Language. This kind of SLA defines how to exlude some users from a service or an asset. So Digital Rights Management (DRM) is a special case of an SLA. The main idea is to manage rights over services in the internet. The scope of ODRL is to complement analogue right standards by providing digital equivalents. Therefore ODRL provides the mechanism to express digital rights policies. ODRL is not developed by a unique work group, moreover the specifications and models are elaborate by different groups over the world [9].

4.2.7.2 Language Description

In Figure 4.9 assets, rights and parties are the three main core entities. The assets are the physical or digital content the right management is talking about. The rights contain the permissions with its underlaying constraints, requirements and conditions for the assets. And finally the parties contain the end user and the right holder. They can be humans, organisations or defined roles. Followed by the definition of these core entities the remaining parts of the Figure 4.9 are going to be defined [9]. ODRL defines all this parts clearly in detail. For this seminar work it's not appropriate to explain this in detail.

4.2.7.3 Agreement Structure

The ODRL Agreement is defined by two XML schemas. One is for defining the expression language and the other for the data dictionary. The expression language contains the elements which are defined in the data dictionary in detail. For example in the language expression there is a element defined, so in the data dictionary schema is defined what this element means and describes. There is no need for natural language, a framework helps developing and using the XML generated agreement [9].

In addition there is an agreement on the side of the publisher of a service. In this policy the rights and usage of the asset are defined, *e.g.* how many copies someone can make. If

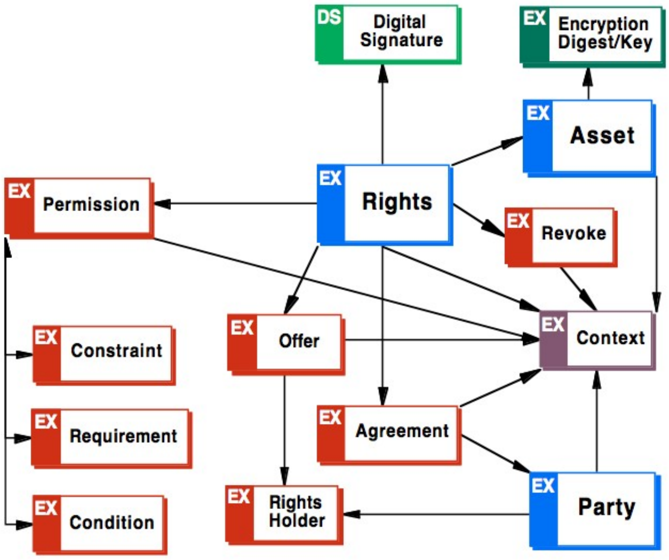

Figure 4.9: ODRL Foundation Model

the user agrees this policy there will be generated an other agreement with the accept of the end-user [9].

4.2.7.4 Development Status

There are no documents available wherever is described if ODRL is still developed at the moment. The language is also developed for extensibility. This means it is possible to define new data dictionaries with creating a new XML Schema that imports the expression language schema as described above [9].

4.2.8 Summary and Conclusion

All of the languages described above have the goal to express Service Level Agreements. They try to formalize and standardize the way of expressing SLA's. In earlier times SLAs were mostly written in natural languages, this gave more freedom to express the agreement but it was not possible to precisely define the parameters and operations which must be guaranteed and monitored. Some of the languages are supporting monitoring in an automated manner, but we want to distinguish between the languages which fully describe the monitoring part of an SLA and the ones supports monitoring without specifying it. As an example the specification of WS-Agreement is a complete framework for monitoring and expressing SLAs with focus on the monitoring part. In contrast SLAng is not supporting the monitoring part as much as the latter. The focus of SLAng is more on the exact expressing of an SLA and only provides certain possibilities to monitor the agreements.

Considering the mentioned languages we can build two groups. The first one contains SLAng, WSLA, WSOL and WSPL and the other ODRL-S and WS-Agreement. Group one refers to languages which do not focus on the monitoring part and group two fully includes it in their specifications.

4.3 Comparison of Standards and Models

After the introduction of the different languages we now compare those Standards and Models using several criteria in a table.

4.3.1 Criteria

After taking into account that the languages are not exactly designed for the same purpose, we took the following criteria to show the disparity and the common parts in the compared Standards.

The *origin* shows who has developed the language. In the second criteria is the *language* as itself, in other words just how the language is expressed.After the origin, the *extensibility* is another interesting point to compare. Here you see how you can extent the language and add new additional agreements. In the next part we investigate the *application*, what this Standard is used for. The *violation control* follows in this list. It covers a part of the monitoring. How it is possible to detect violations of the agreed contract and how to treat them. In the last point we show how these languages are *implemented*. With which tools and codes you can use the Standard.

4.3.2 Table of Comparison

Below, the two tables 1.1 and 1.2 are presented. They aim to compare the different languages. The comparison is done using the mentioned criterias.

Table 4.1: Comparison of SLAng, WSLA and ODRL-S

	SLAng	WSLA	ODRL-s
Origin	Research group at University College London	IBM Development Team	W3C DRM Workshop
Extensibility	The XML structure and the existing semantics allows to extend the language in many different ways	The core WSLA Language is defined abstract. This allows to extend the language in several ways using the abstract syntax.	The abstract defined expression language allows to create new schemas with new parameters. But it is restricted to 6 place holders which can be used.
Language	XML based schema	XML core language with three main parts	Contains of XML based expression language and the XML based data dictionary elements
Application	Horizontal and vertical service agreements around Business processes	Used for Web Services	Used for Digital Rights Management(DRM) to protect and use private digital medias and assets.
Violation control	SLAng uses syntactic and semantic elements like abstract classes to handle violation events	No consideration of monitorability or error is given in the specification. WSLA includes no explicit support for assigning financial penalties	No violation control
Implementations	Eclipse Plugin to produce an SLA contract [14]	WSLA Compliance Monitor embedded in the Web Service Toolkit from IBM [19]	There exists at least one Open Source Implementation which we will not declare especially.

Table 4.2: Comparison of WSPL, WS-Agreement and WSOL

	WSPL	WS-Agreement	WSOL
Origin	Sun Microsystem laboratories	GRAAP Working Group open grid form	Carleton University Canada
Extensibility	Because o the simple structure of policy file that consist of a set of rules it is possible to extend the rule set in any direction	Similar to WSLA the underlaying xml schema is extendable	The offering aspect is the iimport. Quality measurements are defined externally therefore it's not possible to add metrics but to add post cost structures.
Language	It is a strict subset of the XACML	XML based(w3c standard)	XML language. Would be powerful if it had a adequate semantic definition. Needs outlined languages and ontologies.
Application	Policies including authorization QoS and service options in Web-Services	Framework with standardized agreement character to create a monitor agreement of any kind of service	formal specification of various constraints and classes of service for Web Services
Violation control	it support a static or dynamic policy negotiation but not a real time monitoring of violation	The WS-agreement framework specifies a fully lifecycle management of agreements including violation and availability control	Monitoring of the offering is included in the language.
Implementations	Open WSPL [17]	Many implementations available, *e.g.* CREMONA, WSAG4J [18]	At the moment there is no specific implementation available. The language is written manually, *e.g.* java XML Parser.

4.3.3 Summary

On the basis of the two tables can be seen that there are two main drivers for the development of the mentioned Standards. On one hand you have the universities and on the other hand the organizations and big companies like Sun Microsystems and IBM. One strong common point all compared languages have, is the XML. It seems that this way to express SLAs in a defined manner is the most efficient. This fact also allows to easily extend and add agreements. The open source characteristic is dominating for the most of the investigated languages and only a few implementations that cost are on the market. An aspect that differs a lot between the languages is the violation control. Some of them do not support the monitoring and others do. And the other aspect that is not the same is the application. You have got some generic languages like SLAng and specialized ones like WSLA that are designed for one direction. A good example for a very specific language is ODRL-S because it is especially designed for Digital Rights Management.

4.4 Implementation and Future

This section is about common business practices, where we have discovered some interesting facts about what is going on the market and what we think the future of SLAs will look like.

4.4.1 Business Practice

So far we have discussed the most common standards and models for SLA's. The idea of this section is to talk a bit about what is actually used by enterprises. By searching the internet you find a lot of small IT businesses, which are offering any kind of IT services. Most of them use a very simple SLA in natural language to define response time, availability of the service, duration and pricing, but the measurement is absolutely not transparent. We have the impression that it is still not profitable for a small business to use SLAs in an automatic manner, because the effort is too big and the technologies and standards are still in development. That's why these agreements are finalized and customized by the Service Provider and the Service Consumer in natural language.

More interesting are the big enterprises and market leaders which have the resources to provide an environment for expressing and monitoring SLAs in an automated manner. In October 2008 Google announced that they extend their 99.9% reliability of GMail to 4 other products, namely Google Calendar, Docs, Sites and Talk [23]. To make the people trust in these values Google argued that they use these services by them self and therefore downtime affects their business as well as the customers. What does 99.9% mean in the sense of Google? Well it sounds like an outstanding availability, but Pingdom pointed out that in fact it is possible to have a downtime worst case scenario more than 21 hours per day, because downtimes of less then 10 minutes are not counted. They mention as well that short downtimes are much more common than long downtime periods for such services. By avoiding these problems Google makes it much simpler to accredite their SLA [24].

Another big player is Amazon. They have announced the same 99.9% reliability on their S3 service, but in 2007 one year before Google [25]. You can get a refund if your S3 service fails (service unavailable or internal error message) and the availability falls under 99.9%. Very disappointing is that they dont say any word about mesuring this reliability, so again no transparency to customers. And if the SLA gets actually violated you have to report and request your refund manually by yourself until 10 business days after the failure occurs [26].

So the actual situation is far away from the research results in the past years. One reason for sure is that maintenance of systems, which manage SLAs, is a big investment. And also lots of business processes have to be restructured or new processes need to be created. Even Google or Amazon are not willing yet to deliver a usable SLA for its customers. Maybe the fear that more transparency can open the field for new competitors. Unfortunately there is not one official accepted standard or implementation yet that would make it possible even for small business to support SLAs in a automated manner.

4.4.2 Future Trends

At the moment the most SLA languages are defined to automatize generating an SLA. In future time the aim is to develop more frameworks which allow to generate SLAs automatically and have a framework, like an engine, to observe the parameters which are set in the agreement. Moreover there is a tendency to a more complex framework which supports the SLA controlling and generation [1].

On the other side leading scientific market wants to reduce the complexity of the agreements. They believe that the companies focus on too detailed and too technical parameters and forget about the rudimentary things and the economic value. For the user it is easier to interpret economic processes with values instead of get information about technical details, *e.g.* reaction time in milliseconds. They also say that it is important to use established industrial standards wherever possible [27].

As it was described above the violation control is not integrated in the most languages. The controlling of the terms and conditions in a automated manner is the aim of many developer. It can be said the more the SLAs are technical the more it is necessary to control them not manually. An SLA in natural language is not quite complicated to control by a user them self, but to control a detailed technical parameter seems to be much more complicated.

4.5 Conclusion

It was very difficult to find usable information about SLAs and their implementations and documentations, because companies normally do not publish them. After our research we think that at the moment there is no proper standard for expressing SLA. There are several approaches that try to meet the requirements. Every language has one good solution for a specific application but there is no model for the overall problem.

All languages we looked at are Open Source. This allows people to collaborate and develop a globally accepted Standard or Model for expressing and monitoring of SLAs.

Another problem is that most of the implemented SLAs are expressed in natural language. The companies will not change to SLAs in a automatic manner until they get enough benefit from changing to the new way of expressing SLAs. Right now it is the cheapest approach to express them in natural language.

Bibliography

[1] Davide Lamanna, James Skene, Wolfang Emmerich: *SLang: A Language for Defining Service Level Agreements*, 9th IEEE Workshop on Future Trends of Distributed Computing Systems (FTDCS2003), IEEE Computer Society, Washington, DC, 28-30 May 2003, pp 100.

[2] *The SLAng Language - Specification*, UCL, 2006.

[3] Hong-Linh Truong, G.R. Gangadharan, Martin Treiber, Schahram Dustdar, Vincenzo D'Andrea: *On Reconciliation of Contractual Concerns of Web Services*, Distributed Systems Group, Vienna University of Technology, 2008.

[4] *Service Level Agreement, English language*, Available at: `http://en.wikipedia.org/wiki/Service_level_agreement`, Last visited on February 2010.

[5] *RESERVOIR Project Website: Service Manager Scientific Report*, Available at: `http://www.reservoir-fp7.eu/fileadmin/reservoir/delivarables/A4_ServiceManager_ScientificReport_V1.0.pdf`, Last visited on February 2010.

[6] *IBM, Web Service Level Agreement (WSLA) Project: WSLA specification*, Available at: `http://www.research.ibm.com/wsla/WSLASpecV1-20030128.pdf`, Last visited on February 2010.

[7] *WSLA: Web Service Level Agreement*, Available at: `http://www.wsmo.org/papers/presentations/WSLA_04.01.2006.pdf`, Last visited on March 2010.

[8] Ian Toma, Douglas Foxvog, Michael Jaeger: *Modeling QoS characteristics in WSMO*, Available at: `http://elib.dlr.de/44915/1/p42-toma.pdf`, Last visited on March 2010.

[9] Renato Iannella: *Open Digital Rights Language (ODRL)*, Available at: `http://odrl.net/1.1/ODRL-11.pdf`, Last visited on March 2010.

[10] *Open Grid Forum, GRAAP Working Group: WS-Agreement specification*, Available at: `http://www.ogf.org/documents/GFD.107.pdf`, Last visited on March 2010.

[11] Anne H. Anderson: *A Introduction to the Web Services Policy Language (WSPL)*, Sun Microsystems Laboratories. Available at: `http://research.sun.com/projects/xacml/Policy2004.pdf`, Last visited on March 2010.

[12] Vladimir Tosic, Bernard Pagurek, Kruti Patel: *WSOL - A Language for the Formal Specification of Various Constraints and Classes of Service for Web Services*, Research Report OCIECE-02-06, Department of Systems and Computer Engineering, Carleton University, November 2002.

[13] *Precise Service Level Agreements*, Available at: `http://www.cs.ucl.ac.uk/staff/W.Emmerich/publications/ICSE2004/icse04.pdf`, Last visited on March 2010.

[14] *The SLAng SLA Language*, Available at: `http://uclslang.sourceforge.net`, Last visited on March 2010.

[15] *Checkliste SLA*, Available at: `http://www.rwi.uzh.ch/oe/zik/archiv/ChecklisteSLA.pdf`, Last visited on March 2010.

[16] *Language support for service-level agreements for application-service provision*, Available at: `http://eprints.ucl.ac.uk/5607/1/5607.pdf`, Last visited on March 2010.

[17] *Open WSPL*, Available at: `http://sourceforge.net/projects/openwspl`, Last visited on April 2010.

[18] *WSAG4J - WS-Agreement framework for Java*, Available at:`http://packcs-e0.scai.fraunhofer.de/wsag4j`, Last visited on April 2010.

[19] *Web Service Level Agreements Project - WSLA Compliance Monitor*, Available at: `http://www.research.ibm.com/wsla/implementation.html`, Last visited on April 2010.

[20] *ODRL - Wikipedia*, Available at: `http://en.wikipedia.org/wiki/ODRL`, Last visited on April 2010.

[21] *List of WS-Agreement Implementations, English language*, Available at: `https://forge.gridforum.org/sf/wiki/do/viewPage/projects.graap-wg/wiki/Implementations`, Last visited on April 2010.

[22] *eXtensible Access Control Markup Language, English language*, Available at: `http://www.oasis-open.org/committees/xacml/`, Last visited on April 2010.

[23] *Official Google Blog - What We Learned From 1 Million Businesses In The Cloud*, Available at: `http://googleblog.blogspot.com/2008/10/what-we-learned-from-1-million.html`, Last visited on April 2010.

[24] *Pingdom - Google Apps SLA loophole allows for major downtime without consequences*, Available at: `http://royal.pingdom.com/2008/12/04/google-apps-sla-loophole-allows-for-major-downtime-without-consequences/`, Last visited on April 2010.

[25] *ZDNET - Amazon S3 service promises 99.9 percent uptime*, Available at: `http://blogs.zdnet.com/BTL/?p=6554`, Last visited on April 2010.

[26] *Amazon S3 SLA*, Available at: `http://aws.amazon.com/de/s3-sla/`, Last visited on April 2010.

[27] *Nur einfache SLAs sind gute SLAs*, Available at: http://www.computerwoche.de/management/it-services/587629/, Last visited on April 2010.

Chapter 5

Opportunities and Threats of New Technologies to P2P Networks

Michael Ammann, Thomas Kaul, Nicolà Parpan

Peer-to-peer (P2P) networks allow users to share information without the necessary presence of a central server or provider. The adoption of new Internet technologies will evolve the way users and applications interact. The impact on P2P networks by adopting new technologies, such as IPv6 and the Fiber to the Home (FTTH) medium, is presented. IPv6, as the successor to IPv4, offers some new economic opportunities to P2P networks, like a larger address space, Anycast or Mobile IPv6. The extension of optical fiber to the end-user's home with FTTH will lead to a significant performance improvement for P2P networks, caused by larger bandwidth and higher upstream capacities. In addition, threats of these new technologies to P2P networks are considered by analyzing the lack of pressure to adopt IPv6, the deficient availability of corresponding IPv6 products, as well as the fact that even with FTTH, Client/Server systems will still be faster.

Contents

5.1 Introduction

Peer-to-Peer systems are very popular on the Internet. They allow users to share information, such as large files, audio or video streams, without the need of a central server. This paper examines economic opportunities and threats of new technologies to Peer-to-Peer networks. The authors chose among many technologies two new and promising ones with a considerable impact on Peer-to-Peer networks: The Internet Protocol version 6 and the Fiber to the Home medium.

This paper is structured as follows: Section 5.2 introduces the topic of Peer-to-Peer systems and demonstrates advantages, disadvantages, popular applications and protocols from the past. In section 5.3, the two selected technologies, Internet Protocol version 6 and Fiber to the Home, are technically described. Within the Internet Protocol part the following aspects are discussed: Addressing and routing, Network Address Translation, security and the support of mobile devices. Within Fiber to the Home there are topics such as functionality, different architectures, technologies and standards. The next section 5.4 analyzes economic opportunities separately on both technologies. Section 5.5 discusses threats of the adoption of such new technologies. Finally, section 5.6 provides conclusions, discussions and future trends.

5.2 P2P Essentials

A **Peer-to-Peer** (P2P) system is a network that benefits from an end-to-end communication concept in the Internet. It allows users to share information, without the need of a central server or provider. The architecture consists of a distributed network. There are direct connections between each node. P2P is well-established for sharing information, such as large files, audio and video streams, without the presence of a paid provider, while being fault-tolerant and scalable. On the contrary, a **Client/Server** (C/S) architecture uses a central server for storing and publishing information.

A peer is defined as a node that is actively participating in the overlay, *e.g.* a PC running Napster. It acts as a content provider and requester, as well as a router in the overlay network. A peer is identifiable by a unique ID. [2]

A P2P system has the following characteristics:

- **Symmetry and equality:** Each peer can act as a client, as well as a server and all peers are coequal.
- **Decentralization:** There is no central coordination. Hence there is no central node that coordinates actions between the peers. No peer knows the whole system (no global view). Each peer only knows the peers it interacts with.
- **Self-organization:** The configuration of the whole system emerges from local interactions between the peers.

- **Autonomy:** Peers are autonomous in their behavior and decisions.
- **Reliability:** The overlay of the network is reliable up to a certain point.
- **Availability:** All the data that is saved in the system has to be available. This means that independent of connection losses, unknown peers, distributed storage, *etc.* all the data must be accessible up to a high degree of availability.

P2P systems have both advantages and disadvantages in comparison to standard C/S architectures, depending on the C/S type, as shown in the next subsections. [9]

5.2.1 Advantages

Extensibility: P2P systems can easily be extended. It is easy to add new resources to an existing system. Adding new devices is way simpler than in a C/S architecture because in the latter, network configurations are required.

Fault-tolerance: P2P systems are quite fault-tolerant. Even when some peers are defect or lose connection, it does not significantly influence the system. On the contrary: When a server breaks down, a C/S system gets useless.

Scalability: Scalability (how large can a system grow) is often promoted as a key advantage of decentralized systems over centralized ones. Since peers also serve, adding a new server is not necessary.

5.2.2 Disadvantages

Manageability: How hard is it to keep the system working? Complex systems like P2P are harder to manage because every peer requires updating, repairing and logging. In most C/S, only the server needs this maintenance.

Information coherence: How authoritative is information? Non-repudiation, auditability, and consistency are particular aspects of information coherence. In centralized systems like C/S, all information is in one place, the server, therefore the system is fully coherent. In contrary, the widespread nature of decentralized networks causes the system difficult to manage and tends the data to be never fully authoritative. Peers are unreliable.

Security: Security covers a variety of topics, such as preventing people from taking over the system, injecting bad information, or using the system for a purpose other than which the owners intend. In both P2P and C/S architectures, security is an important topic that is not easy to handle. But in P2P networks it is even harder to prevent the distribution of malicious data due to the following facts: A bad node can easily join the network, create new identities and can lie or refuse to provide services.

5.2.3 Applications and Protocols

Listed below there are some important P2P protocols and applications in a chronological order.

Napster was a popular music exchange system established in 1998 by Shawn Fanning. It had a revolutionary effect because of its architecture: After searching for a song, the client connects to other clients and exchanges the data directly. [1]

XMPP (Extensible Messaging and Presence Protocol) is an open protocol based on XML for instant messaging. It has been developed in January 1999 by Jabber, an open source community. In 2002 the Internet Engineering Task Force (IETF) has formed a working group to formalize the core protocols as an IETF instant messaging and presence technology. [3]

SETI@home (Search for extraterrestrial intelligence at home) fascinated the Internet community since May 1999. It is a volunteer computing project using computers with Internet connection. Its goal is to analyze radio signals and search for signs of extra terrestrial intelligence. The project is hosted at the University of California, Berkeley, in the United States. [1]

Another system is called **Freenet**, initially released in March 2000. "Several years before the peer-to-peer mania, University of Edinburgh researcher Ian Clarke started to create an elegant, simple and symmetric file exchange system that has proven to be among the purest of current models for peer-to-peer systems." [1] In Freenet there is no difference between client and server, and there is no centralization. A key feature is the strong protection of anonymity. After uploading a file, the user is able to shut down his node because the content is stored in the network and there are no more dependencies between the user and the uploaded file. [1]

BitTorrent is a P2P file sharing protocol designed for sharing huge volumes of data. It was released in July 2001. By contrast to other file sharing technologies, BitTorrent does not rely on a global network but it establishes a distribution network for each file. [4]

Skype, established in August 2003, is one of the most widely used Internet phone applications. Calls within the Skype network are free, while calls to landline phones or mobile phones are chargeable. The software is using P2P technology and is also popular for instant messaging, file transferring and video conferencing. [5]

5.3 New Technologies

In this chapter, two new technologies, which have an influence on P2P systems, are introduced: Internet Protocol version 6 and Fiber to the Home. While Internet Protocol version 6 is the successor of the common Internet Protocol version 4 (IPv4), Fiber to the Home is concerned with the upgrade on the last mile of telecommunication networks with optical fiber cables.

5.3.1 IPv6

Internet Protocol version 6 (IPv6) was designed as the successor to IPv4, which was first used in the Internet. One major impulse for the development of IPv6 was the lack of address space in IPv4 [8]. IPv6 was defined as a standard in December 1998 by the IETF.

IPv6 has an enormous enhanced address space compared to IPv4. The new Internet Protocol increases the Internet Protocol address (IP address) size from 32 to 128 bits. Therefore the address space supports 2^{128} (about 3.4×10^{38}) addresses instead of an upper limit of 2^{32} (about 4 billion) unique addresses. In the early design stages of the Internet, this amount of addresses seemed to be sufficient. But in just 25 years, the Internet grew from an experiment with a few dozen networks into a worldwide network with hundreds of millions of clients. [6]

Another change in IPv6 can be found in the simplification of the header format (Figure 5.1). Some fields of the header in IPv6 have been eliminated or made optional. This leads to a reduction in common-case processing cost of packet handling and a limit of the bandwidth cost. IPv6 has an improved support for extensions and options with a better flexibility for future options [7]. The new protocol also has the capability of flow labeling. This is "added to enable the labeling of packets belonging to particular traffic 'flows' for which the sender requests special handling, such as non-default quality of service or 'real-time' service." [7] In IPv6 there are also extensions for authentication, data integrity and optional data confidentiality.

5.3.1.1 Addressing and Routing

IPv4 defines three classes of addresses: Unicast, Broadcast and Multicast addresses. In IPv6, Broadcast addresses are not in use anymore, due to the fact that they cause scalability problems in most networks. Multicast addresses are used instead. There are three classes of addresses in IPv6: Unicast, Multicast and Anycast. A Unicast address can identify an interface of an IPv6 node. The packet which was sent to a Unicast address, is delivered to the interface identified by that address. A Multicast address encompasses multiple IPv6 interfaces. A packet sent to a Multicast address is handled by all members of the Multicast group. An Anycast address is assigned to multiple interfaces (mostly on multiple nodes). A packet sent to an Anycast address is delivered to only one of these interfaces, usually to the closest one with that address. These differences are illustrated in Figure 5.2.

Addresses in IPv6 are assigned to interfaces (as in IPv4), but not to nodes such as in the Open System Interconnection Reference Model (OSI model). Therefore every interface of a node needs at least one Unicast address. As mentioned before, a single interface can be assigned to more than one IPv6 address of any type.

An IPv6 address is structured in the following way: Global Prefix, Subnet ID and Interface ID. The global routing prefix is used to recognize a special address (for example Multicast) or a range of addresses assigned to a site. The Subnet ID is used to identify a link within

IPv6 Header Format

```
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|Version| Traffic Class |           Flow Label                  |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|         Payload Length        |  Next Header  |   Hop Limit   |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
+                         Source Address                        +
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
+                      Destination Address                      +
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+

Version              4-bit Internet Protocol version number = 6.

Traffic Class        8-bit traffic class field.

Flow Label           20-bit flow label.

Payload Length       16-bit unsigned integer. Length of the IPv6
                     payload, i.e., the rest of the packet
                     following this IPv6 header, in octets.

Next Header          8 bit selector. Identifies the type of
                     header immediately following the IPv6
                     header. Uses the same values as the IPv4
                     Protocol field.

Hop Limit            8-bit unsigned integer. Decremented by 1 by
                     each node that forwards the packet. The
                     packet is discarded if Hop Limit is
                     decremented to zero.

Source Address       128-bit address of the originator of the
                     packet.

Destination Address  128-bit address of the intended recipient
                     of the packet (possibly not the ultimate
                     recipient, if a Routing header is present).
```

Figure 5.1: IPv6 Header Format [7]

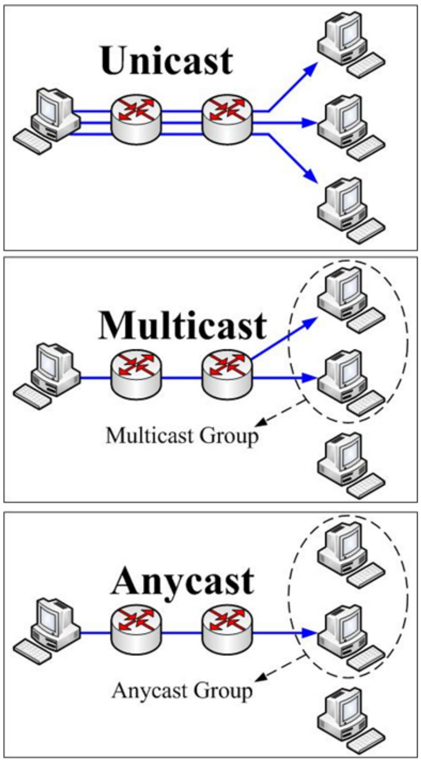

Figure 5.2: Differences in Unicast, Multicast and Anycast [36]

a site and is connected with one link. The interface ID is applied to recognize an interface on a link. It has to be unique on that link. [8]

Scalability issues were the reason for improvements in Internet routing. Some of those existing solutions have been designed for use with IPv6. Although IPv4 continues to hold up fine under considerable pressures of expansion, the routing tables of the Internet have been growing at an explosive rate. These are the backbone routers whose routing tables must reply routing information for every connected network in the world. Some experts have suggested that more than any address shortage, the size of the routing tables will finally drive the acceptance of IPv6. It is not clear that IPv6 will solve these problems because routing has not been improved in IPv6. Both protocols are similar. [6][30]

5.3.1.2 Network Address Translation

The translation of an IP address used within one network to a different IP address known within another network is the process called Network Address Translation (NAT). It is mostly used to overcome the limitations of the small address space of IPv4. For both security as well as network masquerading, which is a technique to hide an entire address space by mapping a large network onto a small IP address space. For example corporate networks use IPv4 addresses for the private network and a NAT router which translates these private addresses to a single public address. [6]

Since IPv6 is strongly upcoming, there was an urgent need that IPv6 hosts were able to communicate with the global Internet, which currently has majority of IPv4 hosts. Simply stated, IPv4 and IPv6 nodes need to coexist and communicate [29]. NAT – Protocol Translation (NAT-PT) can provide this functionality. NAT-PT allows native IPv6 hosts and applications to communicate with native IPv4 hosts and applications, and *vice-versa*. A NAT-PT device resides at the boundary between an IPv6 and IPv4 network. It uses a pool of IPv4 addresses for assigning to IPv6 nodes dynamically, and this assignment is finished when sessions are initiated across IPv4-IPv6 boundaries. [31]

5.3.1.3 Security

Security is an important issue in the Internet due to the fact that it is more globally than in the early days and therefore accessible for nearly everyone. The original designers of the Internet structure based on IPv4 did not have to worry much about security because the Internet was a trusted network for a reliable group of people. To keep up with the growing threats, a standardized and mandatory security protocol was needed. IPv6 addresses several of the problems of IPv4, including the lack of security.

Since IPv4 is an end-to-end model, it relies on the end-nodes to provide and handle security during the communications. IPv4 is hence susceptible to: [14]

Denial of Service (DoS): DoS is an attempt to make a computer resource unavailable to its intended users.

Distributed Denial of Service (DDoS): A DDoS occurs when several systems flood the bandwidth or resources of a targeted system. Usually the target is one or more web servers.

Malicious code distribution: Viruses and worms can use compromised hosts to infect remote systems. This distribution is aided by the small address space of IPv4.

Man-in-the-middle attacks: Due to the lack of authentication mechanisms in IPv4, an attacker is able to read, insert and modify at will messages between two hosts.

Port scanning: In this type of attack, a whole section of a network is scanned to find potential targets with open services. Open ports can be used to exploit the specific hosts further. The small address space of IPv4 makes it easy to scan whole networks in very little time.

Address Resolution Protocol (ARP) poison: The purpose of these attacks is to send fake ARP messages to a network. The aim is to associate the attackers Media Access Control (MAC) address with the IP address of another node. Any traffic meant for that IP address would be mistakenly sent to the attacker instead.

However, many techniques have been developed to overcome some of the IPv4 security issues. NAT and Network Address Port Translation (NAPT) were originally designed to preserve against the fast depleting IPv4 address space, but these techniques provide also a certain degree of protection against the threats mentioned before [15]. IPv6 deals with most of these issues, but does not solve them completely. Another milestone was achieved with the introduction of the IP Security Protocol (IPSec). Although its implementation is optional in IPv4, it is a mandatory part of IPv6.

The IPSec architecture has been defined and standardized by the IETF in RFC 2401 in 1998 [11]. It provides a security architecture for the Internet Protocol (IPv4 and IPv6), not a security architecture for the Internet. An important aspect of IPSec is that it provides an open standard for building security into the network layer, as well as the possibility to create Virtual Private Networks (VPN). VPNs are capable of securely carrying data across the Internet.

The architecture of IPSec includes six elements as shown in Figure 5.3. The IP Encapsulating Security Payload (ESP) and the Authentication Header (AH) are defined as header extensions for the IPv6 protocol or header options in case of IPv4. As their names may imply, the ESP provides an encryption mechanism for authentication, data integrity and confidentiality where as the AH provides an authentication mechanism for authentication and data integrity [6]. Additionally, IPSec provides functionality for protocol negotiation and key exchange management (Internet Key Exchange IKE). This suite enables the mechanisms needed to establish and arrange security parameters between endpoints. It also keeps track of all agreements to ensure secure communication at all times. [14]

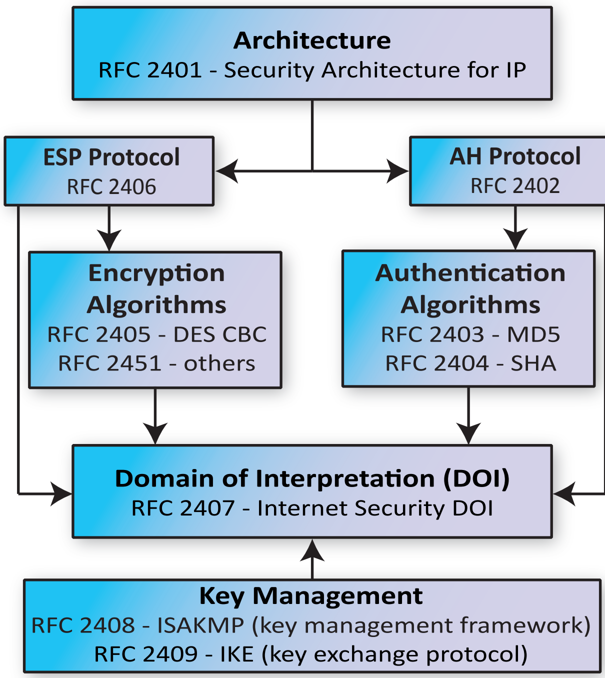

Figure 5.3: IPSec framework [38]

In the following, the process of a secure communication under IPv6 is described. First of all, communicating partners need to agree on a common set of information – a key, authentication or encryption algorithm and some parameters used for one of these algorithms – before they can use the security elements of IPv6. This agreement is called a Security Association (SA) and it is used in every secure communication between certain participants. There exist two types:

Transport mode defines the SA between two end systems for either encryption or authentication for the payload contained in all IP packets related in this communication.

Tunnel mode defines the SA between two security gateways, which surround the IP packet with an outer IP packet (Wrapper). This will allow to use encryption or authentication to the whole inner IP packet.

All defined SAs are stored in an Security Association Database (SAD). They are uniquely identified by three values: The Security Parameter Index (SPI), an IP destination address and a security protocol identifier either AH or ESP. For each security service a SA is required. For example, two communication partners want to both encrypt and authenticate a two-way connection. Therefore a total of four SAs are required, one for each of the two security features in each of two directions. [8]

5.3.1.4 Support Mobile Devices

If a mobile device departs from its home network, the new network will identify its IP address as foreign and will deny the service. The original gateway from its home network is not anymore valid and a session for communication will not be established. This scenario can happen very often because the visited network does change again and again. For instance, if you are on the way in a city using a mobile device.

In IPv4 a mechanism named Mobile IP (RFC 3344) was developed to handle this difficulty. Mobile IP introduces the following entities. [13]

Mobile node is a host or router that changes its point of attachment from one network or subnetwork to another without changing its IP address and that can maintain its original session without interruption even while moving.

Home agent is a router on a mobile node's home network. Packets sent to and received from the mobile node are tunneled through the home agent. The home agent also maintains the locations of the mobile node.

Foreign agent is a router that provides routing services to a mobile node during its visit on a foreign network. The mobile node has to register with the router, which acts as the other end of an IP tunnel.

"The IPv4 mobility protocol offers Agent Discovery services, with foreign and home agents periodically announcing their presence on a network as well as responding to solicitations for agents sent by mobile nodes in the process of connecting to a network." [6]

In IPv6, Mobile IP (RFC 3775) has been extended to handle routing inefficiencies of Mobile IPv4 as well as other drawbacks described later in this section. As long as the mobile nodes are on their home networks, they act like any IPv6 node with a home address. As soon as they move, they need to acquire a care-of address, at which they are reachable.

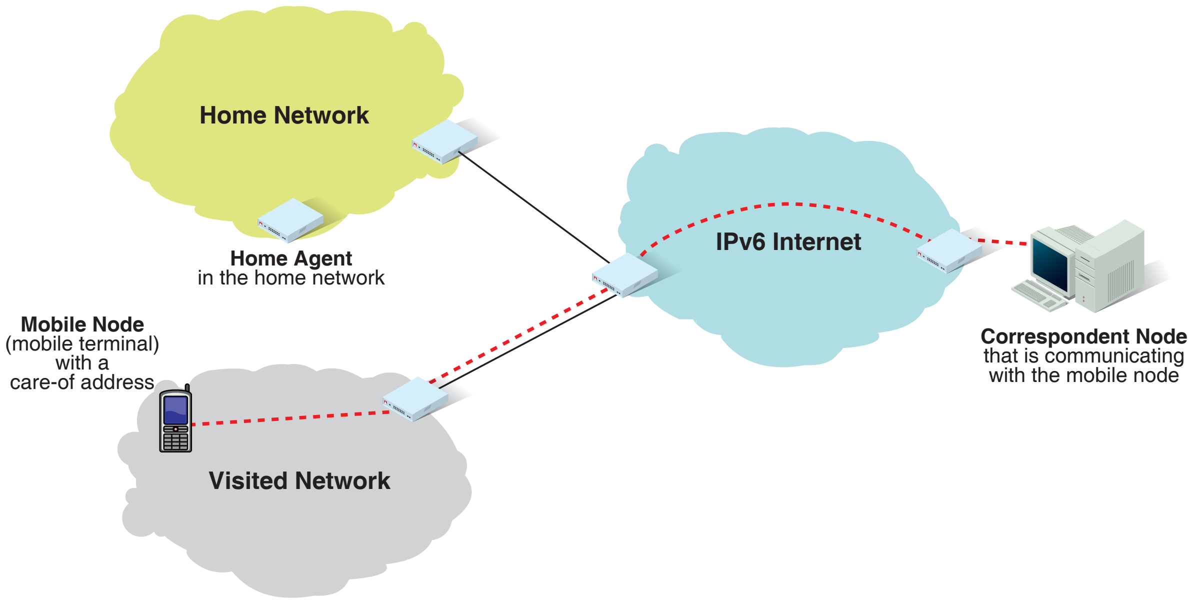

Figure 5.4: Mobile IPv6 scenario [10]

This is done when a mobile node connects to a new network. It gets a local address and binds it together with its home address. Afterwards, an update with the care-of address will be sent to the home agent. This scenario is shown in Figure 5.4. A difference between Mobile IPv4 and Mobile IPv6 is that in the latter a mobile node can notify correspondent nodes, like servers or other mobile nodes about their current locations. [6]

Within this section, the major improvements between Mobile IPv4 and Mobile IPv6 are stated. [6][12]

- Mobile IPv6 operates in any location without any special support required from the local routers. With IPv4 it was necessary to specify routers as foreign agents.
- Route optimization through correspondent nodes lessens the stress on the mobile nodes home network because the operations do not solely run over the home agent anymore.
- In Mobile IPv6, most packets sent to a mobile node, while away from home, are sent using only a routing header, instead of tunneling packets. There is no need to encapsulate/decapsulate packets for tunnels anymore, which improves the processing.
- Mobile IPv6 is decoupled from any particular link layer, as it uses IPv6 Neighbor Discovery instead of ARP. This also improves the robustness of the protocol.

5.3.2 FTTH

Another deployment, which could definitely open new opportunities to P2P networks is the idea of improving the user's physical connection to the Internet. The growing Internet data network plus new online and multimedia services require increasing amounts of bandwidth and better transmission technologies [23]. The existing copper cabling will soon no longer suffice to access communication networks. Because limitations came visible through the last years, there appeared some new technological approaches in this area, transcribed as FTTx.

5.3.2.1 General Description

This is a short introduction to FTTx. With respect to P2P network, the focus lies on FTTH, which is the most dramatic but also most interesting way of implementing FTTx. The generic term "FTTx" stands for "Fiber to the x", where x describes the spot of fiber termination. In case of FTTH, the term stands for Fiber to the Home.

"FTTH is defined as a telecommunication architecture in which a communications path is provided over optical fiber cables extending from the telecommunication operator's switching equipment to (at least) the boundary of the home living space or business office space." [20]

In traditional telecommunication networks, optical fiber is used until several miles away from the end-user's home. The so called "last mile" from *e.g.* a street cabinet to the user's home is often still made of traditional copper cable. With FTTH the optical fiber cable is extended to a full optical path right into the house of each end-user. From there the signal may be forwarded by coaxial cable, twisted pair or wireless media to the different devices.

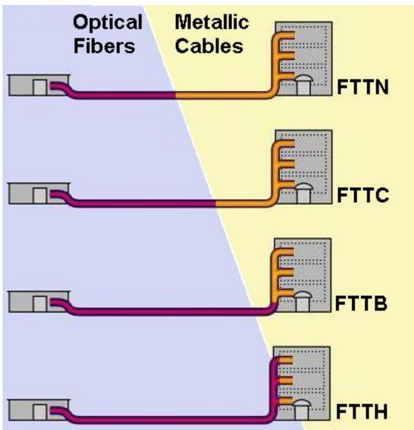

Figure 5.5: Types of FTTx [34]

Besides FTTH, the term FTTx includes some more types of fiber termination terms, like Fiber to the Node (FTTN), Fiber to the Cabinet (FTTC) or Fiber to the Building (FTTB). The difference between all these technologies described with FTTx, lies basically only in the spot where the optical fiber ends and the traditional medium, such as copper cable, begins. A visualization of the different fiber termination spots of these technologies is given in Figure 5.5.

With FTTN, FTTC or FTTB the length of traditional copper wire to the home is shortened, but not down to zero. Somewhere before entering the end-user's home, the optical signal of the fiber is transformed to an electrical signal, which then is transported by a traditional medium. [22] With respect to P2P systems this report will concentrate on FTTH only.

5.3.2.2 How Optical Fiber Works

In order to understand the ideas and deployments on FTTH, it is important to know how optical fiber cables work in general. Optical fiber uses light as a means to transmit data from one location to another. It consists of a light source, *e.g.* a laser or LED, an optical glass fiber as the transmission medium and a detector. The laser generates a pulse of light of a specific frequency, which is detected on the other side by the detector. The detector translates the optical signal into an electrical signal. Thereby a pulse of light normally indicates a '1' and the absence of light a '0'. The electrical signal can then be used by other devices. The lasers signal generating can work at enormous speeds. Commercially available lasers currently reach speeds up to 10 Gbit/s and with the latest technology up to 40 Gbit/s. Recent research in commercial networks has even shown that it is possible to achieve 106 Gbit/s on a single color. Higher speeds are possible but the challenges in reaching these speeds lie in the detector converting light back to electrical pulses. [28]

The optical fiber cable itself consists of multiple layers. In the center there is a glass core which transmits the light. The core is surrounded by a cladding, which is surrounded by a coating or buffer itself. The buffer might be surrounded by a jacket layer made of plastic. An illustration of such a singlemode optical fiber cable is given in Figure 5.6.

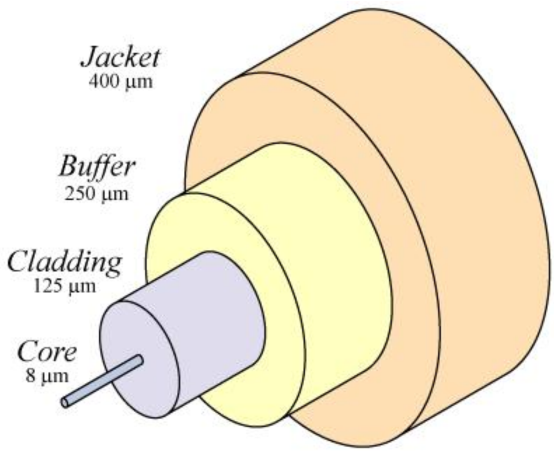

Figure 5.6: optical fiber cable [35]

It is possible to combine multiple fibers to a single cable. Standard cables can carry up to 912 fibers in one single cable. [28]

5.3.2.3 Different FTTH Architectures

There are different ways to combine the fiber cables with all the required equipment to form an efficient FTTH network architecture. Generally an FTTH network constitutes a fiber-based access network, connecting a large number of end-users to a central point known as an access node or point of presence (POP). This POP is further served by a larger metropolitan or urban fiber network, which connects all the access nodes throughout a large municipality or region. The difference between several kinds of FTTH architectures is how the end-users are connected to the POP. [19]

There exist basically two kinds of how the fiber is used for FTTH.

Direct fiber was the first and is still the simplest way to implement FTTH. Direct fiber can be understood as a point-to-point architecture. With that, each fiber goes from the POP direct to each single customer. So each customer has his own dedicated fiber cable from the central to his home. But there is a drawback with direct fiber, which is the obvious extensive amount of required fiber to implement this kind of FTTH architecture. Because of that, some alternative architectures with shared fiber became popular.

Shared fiber is used to reduce the required amount of fiber. It is a kind of point-to-multipoint architecture. Point-to-multipoint architectures for FTTH can either be implemented as a active or as an passive optical network.

Active optical networks (AON) are supported by some electrical power to distribute the signal. This can be done by *e.g.* a router or a switch, mostly using Ethernet technologies.

Passive optical networks (PON) work on the contrary without any electrical support between the POP and the end-users. A PON starts at an optical line terminal (OLT), which is located at the POP. The OLT further connects the whole PON over a single optical fiber cable. This cable reaches to a passive optical splitter, which works without any electrical power. Such a splitter allows the fiber to be split into *e.g.* 16 or 32 additional fiber cables. From there the fiber cables extend to the end-users, where the fiber-based connection ends at a optical network termination (ONT). The ONT at the user's home then transforms the optical signal into an electrical signal, which can be used by other devices.

In a PON the OLT and all the ONTs still require electrical power. Only the splitting works without any electricity, in contrast to a AON where each device requires electrical power.

5.3.2.4 Technologies and Standards

The different FTTH architectures can be implemented by several technologies. Figure 5.7 gives an overview of how the architectures could be implemented by use of optical splitters or Ethernet switches.

Most existing point-to-point FTTH deployments use Ethernet technologies. This works straight forward. At the POP there is an Ethernet switch located, which splits the fiber to dedicated fiber cables laid to each end-user's home.

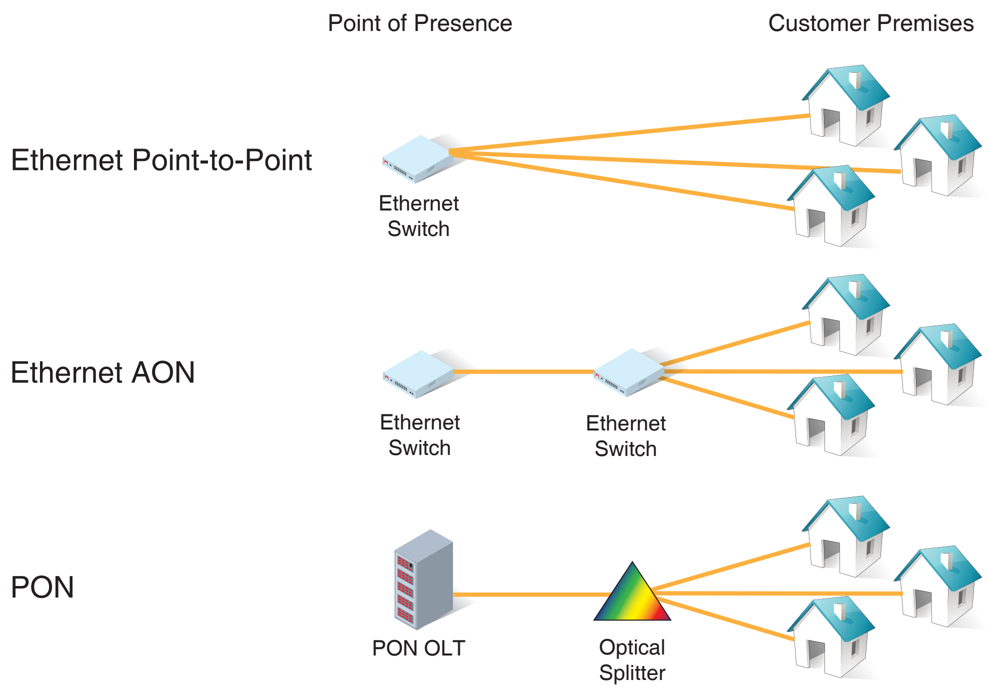

Figure 5.7: FTTH architectures [19]

For that purpose, the Institute of Electrical and Electronics Engineers (IEEE) established the IEEE 802.3ah Ethernet in the First Mile (EFM) Working Group in 2001. This working group created several standards for Fast Ethernet and Gigabit Ethernet. These standards were approved and published in 2004 and were included into the base IEEE 802.3 standard in 2005. [19]

The following list shows the most important standards for Fast and Gigabit Ethernet:

- 100BASE–LX
- 100BASE–BX
- 1000BASE–LX
- 1000BASE–BX

The two end stations communicate with each other in full duplex mode over a single (LX) or dual (BX) strand of optical fiber. 100BASE–X (Fast Ethernet) works with a speed of 100Mbit/s, and 1000BASE–X (Gigabit Ethernet) with 1000Mbit/s. The maximum distance between the communicating stations for all this specifications is 10km. [27]

For AONs as a shared fiber point-to-multipoint FTTH architecture, the technology is similar to the point-to-point implementation. The only difference is that the switch is not located at the POP, but somewhere between the POP and all the end-users. Between the POP and the switch there is only one single fiber cable, and from the location of the switch to the end-users, there is basically a point-to-point architecture with dedicated fiber cables for each user too. Because of this similarity there is often not made any difference between AONs and actual point-to-point architectures.

On the contrary, for PONs there are several different standards of used technologies to implement the point-to-multipoint architecture. The Full Services Access Network (FSAN) Group develops technical specifications, which are then ratified as standards by the International Telecommunications Union (ITU). [19]

The most important standards for PONs are listed below: [19][27]

Broadband PON (BPON) was the first PON standard. It was accepted by ITU in 1999. To carry the data, it uses the Asynchronous Transfer Mode (ATM) protocol, hence BPON is sometimes also called APON. BPON has a downstream link rate of 622Mbit/s and an upstream of 155Mbit/s.

Ethernet PON (EPON) is a standard not introduced by the ITU, but by IEEE in 2004. It uses Ethernet as its native protocol. EPON has a symmetrical down- and upstream of 1Gbit/s. Hence it is also called Gigabit Ethernet PON (GE-PON).

Gigabit PON (GPON) was released in 2003. It relies no longer on the complex ATM protocol but instead on a simpler Generic Encapsulation Method (GEM). GPON has a downstream of 2.5Gbit/s and an upstream 1.25Gbit/s.

Table 5.1: PON standards [25]

	BPON (APON)	EPON (GE-PON)	GPON
Speed - Upstream/Downstream	155/622 Mbit/s	1.0/1.0 Gbit/s	1.25/ 2.5 Gbit/s
Native Protocol	ATM	Ethernet	GEM
Complexity	High	Low	High
Cost	High	Low	Undetermined
Standards Body	ITU-T	IEEE	ITU-T
Standard Complete	Yes, 1995	Yes, 2004	No
Volume Deployment	Yes, in 100'000s	Yes, in 1'000'000s	No
Primary Deployment Area	North America	Asia	Not applicable

Table 5.1 shows an overview about these most important PON standards. There is a lot of deployment going on and new standards will be released in the future. ITU will soon ratify a new standard for GPON, XG-PON, offering 10 Gbit/s downstream and 2.5 Gbit/s upstream. For EPON, there are already products available with 2Gbit/s downstream bit rate. In September 2009 the IEEE ratified a new standard, 10G-EPON, offering 10 Gbit/s symmetric bit rate. Commercial 10G-EPON products will be released in 2010. [19]

Over all FTTH technologies, the PONs are not the dominant architecture yet. In December 2009, 84 percent of all with FTTH connected users in Europe used Ethernet technologies (point-to-point or AON) and only 16 percent really used PONs. [21].

5.3.2.5 State of the Art

“Ninety-nine percent of the Internet’s physical distance has been strung with fiber already.” [22] So fiber has already become the dominant medium for backbone networks but the last mile relies often still on traditional copper cabling.

But the transformation to FTTH has started nearly all over the world. In Switzerland particularly Swisscom has already started to invest in optical plastic fiber networks that can be used inside the home, to accommodate the much higher capacity needed. [22]

The first bigger FTTH construction in Switzerland was built in March 2007 in the city of Basel, where a whole neighborhood out of 190 apartments replaced their traditional copper network access with a modern FTTH network, which supplies optical fiber right into all 190 apartments. [24]

In 2008 already over 4000 Internet customers in Switzerland had an optical fiber access [26]. This number is steadily increasing. At the moment optical fiber cables are laid in all the bigger agglomerations of Switzerland.

In the metropolises of Japan, South Korea or China all the new buildings are throughout realized with FTTH. In the US, FTTH has an annual growth rate of around 100 percent. In Paris, Orange is planning to fit one million households with FTTH and in Germany FTTH has a high double digit annual rate of growth too. [24]

Some market researches have shown that the number of European FTTH households will increase more than five-fold from 2009 to 2013, from four million to over 20 million [23].

But why is FTTH so popular? The reason is simple: FTTH is reliable, scalable and much better than anything else for the same price. FTTH opens totally new dimensions and possibilities for the whole Internet and especially for P2P networks.

5.4 Economic Opportunities

This chapter is dedicated to economic opportunities to P2P networks based on the introduced technologies IPv6 and FTTH.

5.4.1 IPv6

The fundamental design of IPv6 is not a closed system. Because of the open framework approach, IPv6 is an extendible protocol which has nearly no limits in practice. It covers the current needs of functionality and it is open for future requirements. IPv6 is designed to support new and innovative applications such as P2P, but also Sensor Networking, GRID and Ambient Intelligence (AmI). This brings together entrepreneur and researchers from various fields such as computer science, software engineering, electronics and mechanical engineering, design, architecture and social sciences. IPv6 will be the platform of the future. [16]

5.4.1.1 Larger Address Space

The IPv6 address space will be extended from 2^{32} to 2^{128} addresses. This given fact allows a lot of yet existing and new processes to get addresses. It contains for example all the world's billions of mobile phones and computing devices. "It will also include the nearly 200 addressable processes in a typical motor vehicle." [16] Every home can have dozen or hundreds of IP addresses. In the future, every device, which costs more than $10, could have at least one fixed IP address [16]. This also means that NAT is not needed anymore. These conditions may be a great opportunity for P2P technologies. Almost every device can be linked in a P2P network and share information and there are no NATs in use that block incoming or outgoing connection requests.

5.4.1.2 Anycast

While Multicast sessions exist between a distinct source and several destinations (all of them receive the source's signals), Anycast addressing refers to a single source calling a list of Anycast targets. The difference from Multicast to Anycast is that only one destination responds and participates in next transmissions. The other Anycast addresses realize that someone has reacted and do not take part any further. Anycasting has a lot of applications, one of them is the distribution of multimedia and video over the Internet. If a customer requests a video package from a P2P video streaming application, the peer with the nearest location may seed. This can lead to improvement in video quality. It is also possible that a less busy peer with farther distance or an almost overloaded peer with a cheaper path may respond. There are several possible scenarios, most of them were not yet discovered, let alone exploited. [16]

5.4.1.3 Mobile IPv6

In Mobile IPv6 there is a server which is updated regularly with network and gateway information to reach the moving device. Thus the traffic is directly between the mobile device and its communicators but not routed via the home network. This has a positive impact on performance and reliability. Moreover Mobile IPv6 reduces costs. [16]

PEPERS - Mobile Peer-to-Peer Security Infrastructure: This project is an attempt to design, implement and validate a dependable platform with high-level support for the design, development and operational deployment of secure mobile P2P applications for future AmI environments. It is funded by the EU's Framework Programme for Research and Technological Development.

The platform focuses on three major aspects that are used for and support mobile P2P applications (based on the open-source Symbian OS). The definition of advisable security services over suitable protocols, the analysis and design of future platforms and interfaces in mobile devices to contribute security services and the definition of interfaces which will allow secure mobile access to application servers.

The platform intends to rise the number of secure, industry specific P2P mobile applications rapidly. The consequence is to increase the competitiveness of European business.

An example that was developed with the PEPERS platform is a journalism application. It allows reporters and photographers to collaborate on a news story. First of all, a story is assigned to a team consisting of reporters and photographers. Through the P2P nature of the application, they can co-operate, communicate and share information on-the-fly, compile a report (including photos that have just been taken) and send it to the editor. Such applications makes collaborative reporting easier and faster. It also gives an advantage in competition with other newspapers because the faster the news are out, the better. In contrast to IPv4, it is easier to develop P2P mobile applications with IPv6. [37]

5.4.1.4 Various Opportunities

Listed below there are some various opportunities of IPv6 to P2P networks.

Economies of scale: IPv6 allows an increase of the scale of networks. That brings more users with their demands and requirements together. It also simplify extensive deployment of beneficial services that were slow to the IPv4 market.

Direct access to all devices: IPv6 is able to throw out devices which destroy the symmetry (ability to run bidirectional communications) of the Internet. Because of the extended address space every device running with IPv6 can be directly accessed. That opens a new field for innovation: fully distributed services and applications.

New capabilities: "With new capabilities such as simple provisioning mechanisms and protocol extensibility, IPv6 enables communications infrastructures to offer new services that can drive innovation." [30]

Easier market space to enter: IPv6 knowledge is not widespread yet. At the moment there is less competition in IPv6 than in IPv4 environment. This affords smaller companies to innovate and take share of the market. This can be the reason why some firms who want to enter the IP communications market focus on IPv6 rather than contest established IPv4 providers and vendors. [30]

5.4.1.5 Case Study: P2P Inference Engine

In Web 3.0 (also known as Semantic Web) P2P Inference Engines can be running on millions of users' computers. With Semantic Web tools they can produce an information infrastructure a good deal more powerful than the infrastructure that for example Google is currently using. The development in Web 3.0 of P2P Inference Engines merged with standardized domain-specific ontologies is a potential threat to the present search engine model. In principle, all the devices that are connected to the Internet could be used for "all reasoning collaboratively across the different domains of human knowledge, processing and reasoning about heaps of information, deducing answers and deciding truthfulness or falsehood of user-stated or system-generated propositions." [32]

Web 3.0 could bring a move from centralized search engines to P2P Semantic Web Inference Engines, because the P2P community as a collective would have enormously more power, in quality as well as in quantity, than a company like Google can ever have with their server farms. P2P Semantic Web Inference Engines could have a similar impact to Google like P2P file sharing such as BitTorrent to centralized file hosting. [32]

5.4.2 FTTH

Fiber is the future of telecommunication. Generally fiber brings a lot of improvements in contrast to traditional copper cables. The fiber cable itself does not cost much more than traditional kinds of cables but it is much more reliable and the maintenance is very easy. [22]

FTTH could revolutionize the behavior of each and every individual using the Internet. First of all, FTTH brings an extremely higher bandwidth.

The maximum bandwidth of a connection over the physical link depends on the physical characteristics of the link [28]. What makes fiber a winner is its combination of low attenuation of very high frequencies and very low noise, which are the key factors that determine how much and how fast information can be transmitted [22]. Because of the low attenuation and dispersion there are no or only a few repeaters and signal regenerators required, which leads again to lower costs of equipment [28].

FTTH opens totally new dimensions for data transmission, which could benefit the usage of P2P networks dramatically. A significant problem of existing P2P systems is the low average upload capacity on end hosts, caused by the asynchronous bit rates of most traditional broadband technologies.

With FTTH the trend goes towards symmetrical bit rates for up- and downstream. The increase of upstream will be much higher than the increase of downstream. This could be a huge advantage for P2P networks, since the bottleneck for *e.g.* file sharing in a P2P network is the, for the present still very limited, upload capacity.

With *e.g.* GE-PON there are already symmetrical bit rates of around 1Gbit/s possible. This means that the upstream can be up to 2000 times faster than with common ADSL technologies, while the downstream is at least 200 times faster. A comparison between common ADSL and GE-PON is shown in Figure 5.8.

With GPON, which has an upload capacity of around 1.25Gbit/s, the upload works already up to 2500 times faster than with ADSL, the download on the contrary around 500 times faster, as shown in Figure 5.9.

Higher speeds with fiber will be possible in the future, since the limiting factor is the converting of the light signals back into electrical signals which happens on the ONTs. Nowadays there is a lot of research and product improvement going on for such detectors

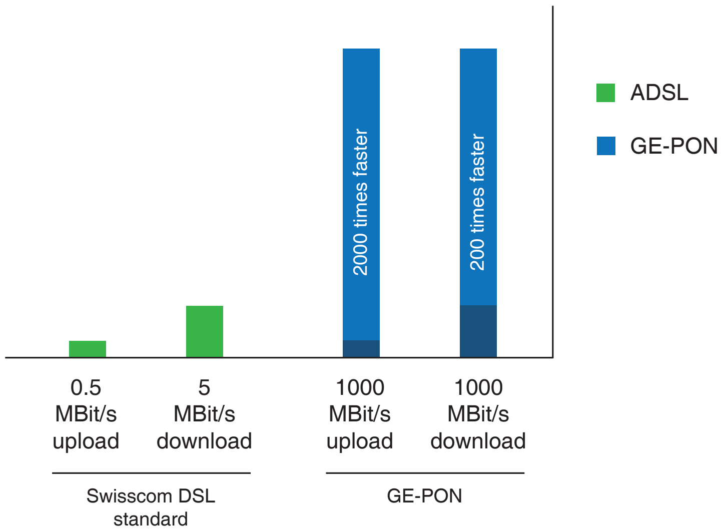

Figure 5.8: DSL - GE-PON comparison

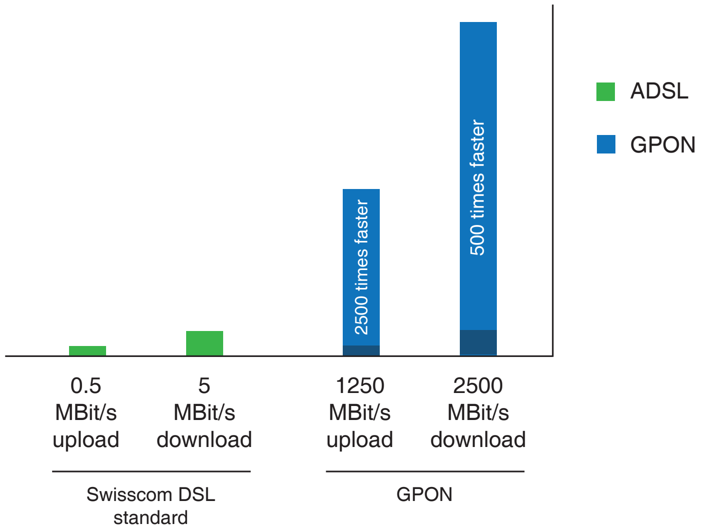

Figure 5.9: DSL - GPON comparison

to reach higher speeds in optical fiber communication. The actual speed of FTTH will also depend on other factors like the distance to the central office and the split ratio to share the fiber.

Since a fiber connection to the end-user's home is very stable and reliable, and the maintenance costs are very low, a P2P network will be easier to be kept maintained with FTTH than today with traditional connections. The stable and reliable connections for relatively low investment costs in combination with a much higher speed performance, caused by the dramatically raised upload speed on end-users, could lead firms and institutions to prefer P2P networks instead of the so far leading C/S networks.

5.5 Threats

This chapter is concerned with possible reasons why the establishment of IPv6 is taking that much time. Potential threats for IPv6 and FTTH will also be treated.

5.5.1 IPv6

In real world IPv6 prevails not that fast. The address assignment for IPv6 has changed from an experimental to a normal operation in July 1999 [17]. More and more Internet Service Provider (ISP) operate in addition to IPv4 also IPv6 on their networks. On the part of end-users IPv6 is not necessarily required, because beside the bigger address space all integral new features of IPv6 has been ported back to IPv4 with more or less success. There is no widely-used application, which only works in an IPv6 environment.

In the meantime a lot of operating systems such as Linux, Mac OS X, Windows Vista or Windows 7 are supporting IPv6. But crucial for a tunnel-free connection is also the support of the firmware for the devices and the router. [39]

By now the spread of IPv6 is still low, but it is growing week by week. It will take more time because there are great differences between various countries (Table 5.2) and the adoption of IPv6 is slower than predicted. [18]

Table 5.2: Connectivity by country [18]

Country	IPv6 penetration
Russia	0.76%
France	0.65%
Ukraine	0.64%
Norway	0.49%
United States	0.45%
...	...
China	0.24%
Japan	0.15%

Why does it take such a long time? What are the challenges of the adoption of IPv6? Here are some problems from an industry perspective:

IPv6 is not a feature: A common perspective can be that IPv6 is seen as a feature, which brings new capabilities. The problem with this perspective is that it highly underrates the effort of integrating IPv6 and it minimizes the potential of an IPv6 environment. IPv6 is not just a feature, it is the entire infrastructure of future IP services and communications. [30]

Search for a definite Return on Investment (ROI): The obvious deficit of ROI and popular applications is nowadays still the central barrier to adoption. But the major goal remains to scale the IP environments to the needs of the Next Generation

Networks (NGN). In this infrastructure role, IPv6 needs a much more complex ROI calculation, if possible at all. [30]

These two misconceptions slow the development of practical challenges in industry.

Lack of pressure to adopt IPv6: Due to the fast depleting IPv4 address space, the call for IPv6 was great. But with the introduction of NAT and Classless Inter-Domain Routing (CIDR), a technique for an efficient usage of the 32-bit address space of IPv4, the immediate need for IPv6 has been postponed. In the last years, many predictions were made when the address space exhaustion of IPv4 will finally come, but the majority of people was still hoping for other workarounds like NAT and CIDR. The newest prediction of the Internet Assigned Numbers Authority (IANA) states that on September 2011 the pool of unallocated addresses will be depleted. So there is no more time to wait to adopt to IPv6. [30][17]

Investment in adoption: To adopt oneself - from the view of an enterprise - to IPv6, significant restructuring is necessary that will cause large investments. So far, for most organizations it is hard to weigh the benefits of IPv6 against the costs that this change will carry along, so they hesitate in adoption. However, business cases show that an early planning was proven to significantly reduce costs. [30]

Availability of products: IPv6-ready products, which means products that are able to make use of the new features containing in the IPv6 protocol, are an important pillar in the advance of IPv6. Global manufacturers have been developing such products for a long time, but localized manufacturers might have just started to introduce IPv6 in their products recently. The more people are using IPv6, the faster it will prevail. [30]

Reluctance of ISP: Due to fixed IP addresses that are coming along with IPv6, there is a lack of incentives for the ISP. They lose the possibility to charge on the distribution of static IP addresses.

Deficiency of experienced people: The number of experts in IPv6 is limited today. Most of the network specialists are not familiar with the new protocol. Training and continuing education is often subordinated to the top costs for adoption. On the other hand, consulting agencies have already been raising their offerings. [30]

5.5.2 FTTH

Beside the opportunities of FTTH to P2P networks, there also some other FTTH-related effects, which could alter the rivalry between C/S and P2P architectures. So FTTH will not only boost P2P networks, it will probably boost traditional C/S networks even more.

For sure FTTH will diminish the performance gap between C/S and P2P networks. But even if the upload speed will increase dramatically, traditional C/S architectures will likely still be faster than P2P, than the download speed will be much higher too with FTTH. This could have positive effects on the use of C/S networks. For P2P the much higher download rate does not have to be that positive, because the bottleneck in performance of P2P networks is still the upload speed. As a consequence the deployment of FTTH will also benefit C/S networks and boost its domination against P2P even further.

A disadvantage of C/S in contrast to P2P is that the connection to the server in a C/S network has to be guaranteed at all time to serve its purpose. In case of P2P, if the connection to a single peer gets lost that would not be so bad, since other peers are still available to compensate the breakdown. Because of that, FTTH, with its higher reliability of fiber connections, will improve the stability of C/S networks probably more significantly than of P2P networks. This could corroborate the preferential treatment of C/S networks in a non insignificant way.

5.6 Summary and Conclusions

Opportunities and threats of new technologies to P2P Networks were presented by analyzing P2P networks, the newest version of the Internet Protocol IPv6 and the transmission technology FTTH. The most important aspects of each have been described in detail, still leaving some space for further work in these topics. Both technologies have been proven interesting, but are not significantly influencing P2P networks on their own yet.

Despite the development of IPv6, a lack of interesting business models have caused this halt of progress of P2P networks and applications. When IPv6 will be established worldwide, there will definitely be a rise in new P2P applications. As a conclusion, it seems that P2P is not boosting IPv6 but IPv6 will boost P2P.

Fiber itself covers over 90 percent of the Internet's physical distance so far, but the last mile is still barely strung with fiber. So the coverage and therefore the use of FTTH is still low. The examination of FTTH led to the conclusion that with immensely higher upstream rates, which are possible by FTTH technologies, the performance of P2P systems will be increased dramatically. Nevertheless, future C/S networks will likely still be faster, since FTTH will also allow much higher downstream rates. But the trend with FTTH goes towards symmetrical bit rates, which could be an important advantage for P2P and its future use.

Bibliography

[1] A. Oram: *Peer-To-Peer: Harnessing the Power of Disruptive Technologies*, O'Reilly & Associates, Sebastopol, CA, 2001.

[2] B. Stiller:*P2P Generations, Past and Future*, Vorlesungsslides, Communication Systems Group, IFI, University of Zurich, 2010.

[3] XMPP Standards Foundation. `http://xmpp.org`, last visited: 13 April 2010.

[4] J. McBrewster, F. Miller, A. Vandome: *BitTorrent (protocol)*, Alphascript Publishing, Mauritius, 2009.

[5] Skype. `http://www.skype.com`, last visited: 20 April 2010.

[6] P. Loshin: *IPv6 - Theory, Protocol and Practice*, Second Edition, Morgan Kaufmann Publishers, San Francisco, 2004.

[7] S. Deering, R. Hinden: *Internet Protocol, Version 6 (IPv6)*, Request For Comments(RFC) 2460, Dezember 1998. `http://www.ietf.org/rfc/rfc2460.txt`, last visited: 16 April 2010.

[8] S. Hagen: *IPv6 Essentials*, O'Reilly Media, Sebastopol, CA, 2002.

[9] N. Minar: *Distributed Systems Topologies: Part 2*, O'Reilly Open P2P, December 2001.

[10] Nokia Networks: *Introducing Mobile IPv6 in 2G and 3G mobile networks*, Nokia Group, Finland, 2001.

[11] R. Atkinson, S. Kent: *Security Architecture for the Internet Protocol* , Request For Comments(RFC) 2401, November 1998. `http://www.ietf.org/rfc/rfc2401.txt`, last visited: 16 April 2010.

[12] J. Arkko, D. Johnson, C. Perkins: *Mobility Support in IPv6*, Request For Comments(RFC) 3775, June 2004. `http://www.ietf.org/rfc/rfc3775.txt`, last visited: 16 April 2010.

[13] C. Perkins: *IP Mobility Support for IPv4*, Request For Comments(RFC) 3344, August 2002. `http://www.ietf.org/rfc/rfc3344.txt`, last visited: 20 May 2010.

[14] Information Technology Protective Security Services: *IPv6-to-IPv4 Transition And Security Issues*, Jalan Gadong, Brunei Darussalam, February 2008. http://www.brucert.org.bn/files/IPv6-to-IPv4%20Transition%20&%20Security%20Issues.pdf

[15] S. Sotillo: *IPv6 Security Issues*, East Carolina University, 2006. http://www.infosecwriters.com/text_resources/pdf/IPv6_SSotillo.pdf

[16] M. Biber: *IPv6 Business Case Scenarios*, Internet Society of Australia. http://www.ipv6.org.au/business.html, last visited: 11 April 2010.

[17] IANA: *Delegation of IPv6 address space*, July 1999. http://www.iana.org/reports/1999/ipv6-announcement.html, last visited: 16 April 2010.

[18] L. Colitti: *Global IPv6 Statistics - Measuring the current state of IPv6 for ordinary users*, RIPE 57, Dubai, October 2008.

[19] FTTH Council: *FTTH Handbook*, Third Edition, March 2010.

[20] FTTH Council: *FTTH Council - Definition of Terms*, August 2006. http://www.ftthcouncil.org/sites/default/files/FTTH_definitions.pdf

[21] IDATE: *FTTH Panorama: EU36 & ME14*, February 2010. http://www.ftthcouncil.eu/documents/studies/IDATE_FTTx_Panorama_Dec_09_v2.pdf

[22] H. Wagter: *Fiber-to-the-X: the economics of last-mile fiber*. http://arstechnica.com/telecom/news/2010/03/fiber-its-not-all-created-equal.ars/, last visited: 11 April 2010.

[23] Reichle & De-Massari: *FTTH: Europe in the Fiber Optic Age*, Media Release, Wetzikon, June 2009. http://www.rdm.com/en/desktopdefault.aspx/tabid-1578, last visited: 16 April 2010.

[24] Swissfibertech: *FTTH weltweit im Vormarsch*. http://www.swissfibertech.ch/content.cfm?nav=1&content=3, last visited: 11 April 2010.

[25] PMC-Sierra: *FTTH Fiber to the Home Overview*. http://www.pmc-sierra.com/ftth-pon/ftth_overview.html, last visited: 20 April 2010.

[26] BAKOM: *Fernmeldestatistik 2008*, Biel, February 2010. http://www.bakom.admin.ch/dokumentation/zahlen/00744/00746/index.html?lang=de, last visited: 11 April 2010.

[27] J. Murphy, D. Nowak: *FTTH: The Overview of Existing Technologies*, Dublin, March 2005. http://www.csi.ucd.ie/Staff/jmurphy/publications/921.pdf

[28] Organisation for Economic Co-operation and Development: *Developments in fibre technologies and investment*, Netherlands, 2008. http://www.oecd.org/dataoecd/49/8/40390735.pdf

[29] Cisco Systems, Inc.: *Cisco Network Address Translation (NAT)*, USA, July 2007. http://www.cisco.com/en/US/prod/collateral/iosswrel/ps8802/ps6969/ps1835/prod_white_paper09186a008011ff51.pdf

[30] P. Grossetete, C. Popoviciu, F. Wettling: *Global IPv6 Strategies: From Business Analysis to Operational Planning*, Cisco Press, Indianapolis, USA, 2008.

[31] P. Srisuresh, G. Tsirtsis: *Network Address Translation - Protocol Translation (NAT-PT)*, Request For Comments(RFC) 2766, February 2000. `http://www.ietf.org/rfc/rfc2766.txt`, last visited 20 May 2010.

[32] M. Fawzi: *The People's Google*, Seattle, July 2006. `http://evolvingtrends.wordpress.com/2006/07/11/p2p-search-the-peoples-google`, last visited: 21 April 2010.

[33] IANA: *IPv4 Address Report*. `http://www.potaroo.net/tools/ipv4/index.html`, last visited: 16 April 2010.

[34] `http://upload.wikimedia.org/wikipedia/commons/thumb/3/32/FTTX.png/450px-FTTX.png`, last visited: 16 April 2010.

[35] `http://upload.wikimedia.org/wikipedia/en/4/4e/Singlemode-optical-fibre.png`, last visited: 27 April 2010.

[36] `http://conf.ncku.edu.tw/research/articles/e/20071123/5.html`, last visited: 20 April 2010.

[37] `http://www.pepers.org/?Page=Project`, last visited: 20 April 2010.

[38] `http://www.lincoln.edu/math/rmyrick/ComputerNetworks/InetReference/141.htm`, last visited: 25 April 2010.

[39] `http://ipv6int.net/systems/index.html`, last visited: 20 May 2010.

Chapter 6

Economics and the Digital Divide

Jeton Memeti, Florim Shabani, Sebastian Müller

The definition of the term Digital Divide has evolved over the last years. Various authors use different definitions, but commonly they agree that Digital Divide refers to the fact that some people have effective access to digital information and technology whereas others have only very limited or no access at all. The access to digital information includes both the physical access as well as the education and skills needed to participate in the digitalized world. Especially developing countries have to struggle with the negative consequences of the Digital Divide which are amongst others the missing competitiveness in economic markets and the lack of possibilities to access information-rich medias like for example the Internet. The main reasons for the Digital Divide are economic factors. Many developing countries cannot afford the necessary investments to build an appropriate infrastructure to guarantee the access to information technologies. In order to measure the Digital Divide, researchers are often using statistics about Internet access or mobile phone penetration rates to compare various countries with each others.

There are reams of initiatives and projects launched by governments, societies, enterprises and individuals to bridge the gap. Some of them were really successful whereas others turned out to be a breakdown. The One Laptop Per Child (OLPC) project and the Wikipedia project are some of the most famous approaches to overcome the Digital Divide. All of the approaches to bridge the Digital Divide which we have analyzed in our work were failures because none of them could really bridge the gap. The main reason for this disappointment is that all the initiatives focused on providing physical access to disadvantaged people. Accompanying actions to teach the people how to use this media were strongly disregarded.

Actually researchers are working on identifying other barriers than the physical access to information technologies which prevent people in developing countries from participating in the digital world.

Contents

6.1 Introduction

The Internet is a lovely invention - everyone who is willing to learn can access the world wide web and easily acquire knowledge. Apart from being willing to learn it requires a computer and the access to the Internet. But in developing countries this fact poses a huge problem. Because of the smaller purchasing power and the marginal earnings the access to a computer and to the Internet is not always possible. There are the costs by the hardware (buying a PC, monitor, modem, etc.), by the Internet (basic fee and charge fee) and of course for the software used with a computer (the operating system, other additional software). The poorer and less educated one is the higher is the probability of not having access to the Information and Communication Technology (ICT) [1]. This leads to the Digital Divide which is a gap between people with access to ICT (mostly in developed countries) and those without access (mostly in developing countries).

The question to response to is if the Digital Divide has been tried to be solved and where the main problems have been, if the approaches were not effective. In Section 6.2 the term "'Digital Divide"' is defined and the actual situation and the emerging problems are described. Furthermore the factors which create and affect the Digital Divide are presented. There exist several approaches to bridge the Digital Divide, namely the OLPC (One Laptop per Child) project, Internet access via mobile phones and Open Source Software. This will be the topic of Section 6.3. Finally in Section 6.4 it will be evaluated if the Digital Divide is solvable at all and if yes how it can be solved.

6.2 Problem statement

In Section 6.2.1 the term "'Digital Divide"' is defined. To comprehend the subject within a greater scheme of things, the actual situation and the emerging problems are explained in Section 6.2.2. The factors which create and affect the Digital Divide, namely economic and educational factors as well as regulatory quality and electricity consumption, are presented in Section 6.2.3.

6.2.1 Definition of Digital Divide

There are terms that stand for the same thing for everybody everywhere at anytime (for example physical formulas, which nonetheless can change over the time). But the more usual case is that a definition of a term is used in different manners depending on the author while the keynote remains more or less the same. And often the meaning of a term shifts over the time. So the term 'Digital Divide' is not always defined in the same way and has changed its meaning over the last years.

In 1996, Amy Harmon, then a journalist at the Los Angeles Times, wrote a story about a couple. The man was spending a lot of time online whereas the wife did not use the computer. This induced a split between the husband and the wife, because she was neglected. Harmon called this a 'Digital Divide' [2]. For Chen and Wellman the 'Digital Divide' is the gap between the existence and the nonexistence of resources to participate in the information era, while the required resources to participate in the information age are telecommunication policies, infrastructures and education [3]. Chinn and Fairlie describe this problem simply as "'international differences in information and communication technology diffusion"' and call it the 'Global Digital Divide' [4]. Global Digital Divide refers more to differences in technology access between countries instead of differences between individuals in a country.

The main issue in the two definitions above is that the access to digital technology causes the problem of Digital Divide. In most cases this is a financial problem, namely the inability to purchase a Personal Computer or to pay for the Internet and to get access to the information era. Some authors go one step further and note that the Digital Divide is not only affected by the physical access to computers and connectivity. Warschauer shows that the additional resources which allow people to use technology well like literacy and education as well as relevant content in diverse languages are at least as important as the physical access to computers [5]. So it's important too, to know how to use a computer instead of just having one. Just as important is the existence of relevant content in the language one understands.

There have been further projects which support this theory. One of them was the "'Hole in the Wall"' experiment [5]. In 2000, the government of New Dehli and an information technology corporation established a project to provide computer access to the city's street children. No teachers were provided, so the children should learn on their own how to use the computer and the Internet. In fact they taught themselves basic computer operations, but rarely used the Internet by reason of lacking content available in their language. They spent almost all their time by drawing with paint or playing computer games and neglected the school by doing so. The "'Hole in the Wall"' project showed that in addition to the physical availability of computers and connectivity other issues like content, language, education or literacy are just as important [5].

Furthermore the term 'Digital Divide' implies a binary criterion: you have access to a computer (the Internet) or you do not. As Cisler points it out, there is a gradation based on different degrees of access to information technology instead of a binary decision criterion [2]. A student at the University of Zurich who can access online material with a highspeed connection at the university has not the same degree of access as an rural citizen in India, who can let print out online material by his neighbour who has an Internet connection at home. Obviously, the student has a higher degree of access.

When talking about the Digital Divide one should always think about the aspects mentioned above. In this paper the term 'Digital Divide' is mostly used to describe the gap between people who have access to computers and the Internet and those who don't.

6.2.2 Actual situation and emerging problems

Information technology is an important source of a country's economic growth and wealth. Röller and Waverman discussed this by comparing the telecommunications infrastructure and the economic development in different countries [7]. Dewan and Kraemer used a more generic approach and explored the correlation between returns from IT investments and GDP (gross domestic product) in different countries [8]. Therefore it is important to examine the Digital Divide and to reduce the gap wherever possible, in order to assure a fair economic competition and to reach a sustainable worldwide economic growth [9]. Although a definitive study is lacking, the empirical literature agrees on the fact that "these differences in technology diffusion may have substantial economic consequences" [4].

The following section examines the occurrence of a Digital Divide in different countries by comparing the access to the Internet.

During the last decades the usage of computer and Internet has increased dramatically. In the beginning of the Internet era, during the early 1990s, less than 1% of the worldwide population had access to the Internet [4]. This increased to 8.1% of the world's population by 2001 [4] and 12.7% by 2004 [10]. As of September 2009 approximately 1.73 billion people worldwide, or 26.6% of the world's population, were using the Internet, according to studies by Miniwatts Marketing Group [6].

Figure 6.1 shows the total sum of worldwide Internet users, ordered by geographic regions, by September 30, 2009. As one can see, most of these users are residing in Asia, followed by Europe and North America.

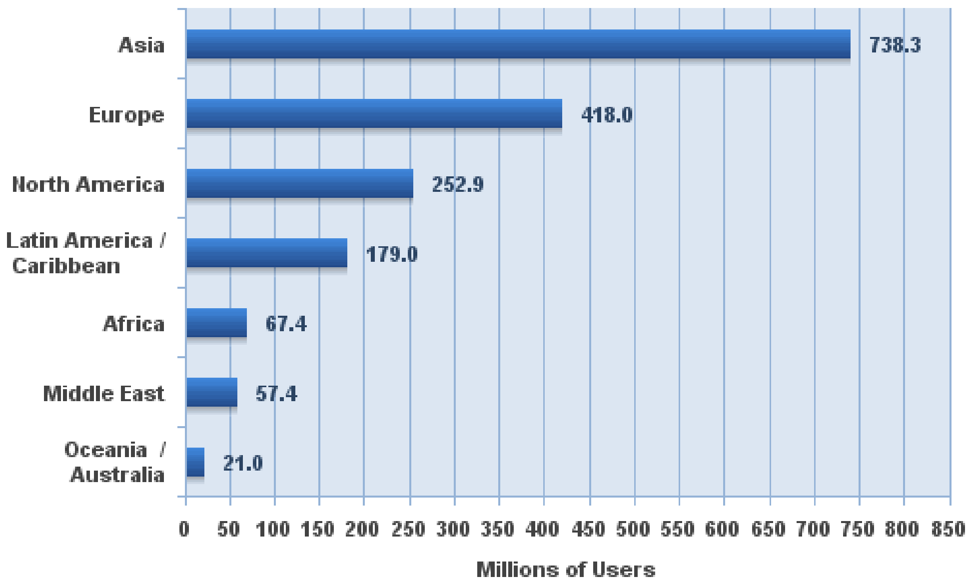

Figure 6.1: Internet users worldwide by geographic regions for september 30, 2009 [6]

This figure implies that Internet access is very common in Asia, Europe and North America and not really widespread in Africa, Middle East and Oceania/Australia. But, in order to analyze a geographical Digital Divide between different countries and continents, total amounts of users aren't really meaningful. It is more significant to analyze the penetration rate, that is, the fraction of the whole population of a region which has access to the Internet.

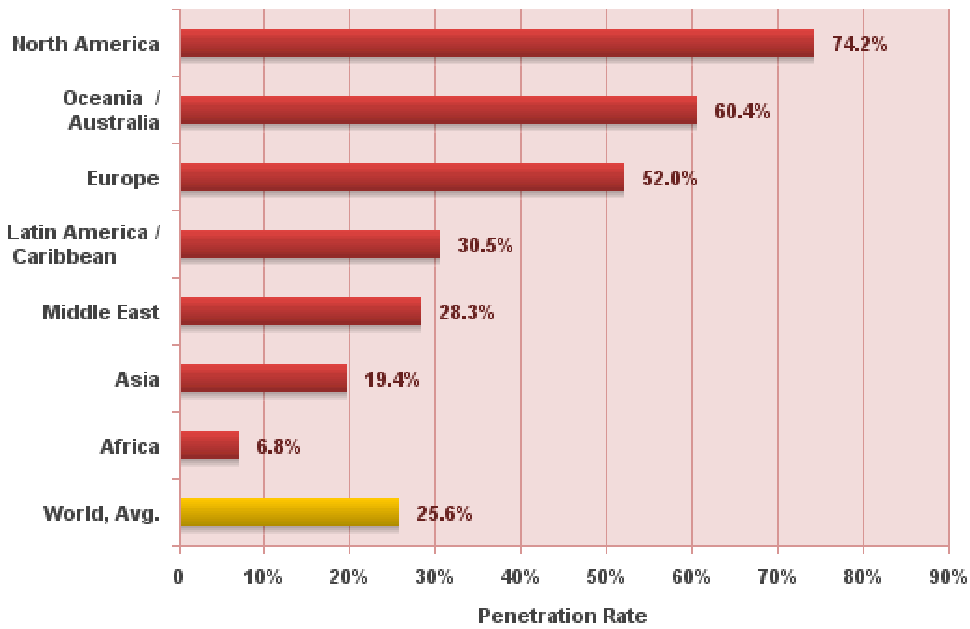

Figure 6.2: Internet users worldwide by geographic regions for September 30, 2009 [6]

A comparison of worldwide Internet penetration rate grouped by geographic regions (Figure 6.2) reveals similar regional relations as in Figure 6.1. There are big differences between different regions and continents, but as in Figure 6.1, one can divide the regions and continents in more or less two groups: one where most of the people have access to the Internet, and the other where most of the people don't have access to the Internet. In North America, Australia and Europe more than the half of the population has access to the Internet, but in Latin America, Middle East and Asia the penetration rate is significantly lower. Africa at the tail end has a very low penetration rate. Only 68 of 1000 Africans have access to the Internet. This rate is almost 11 times smaller than in North America. The penetration rate of Asia is especially remarkable. Although top ranked in the amount of users the penetration rate is far below the world average. This is because in Asia are living almost 57% of the world population [53].

As mentioned in Section 6.2.1, Chen and Wellman define the Digital Divide as "the gap between individuals (and societies) that have the resources to participate in the information era and those that do not" [3]. Bearing in mind this definition, a comparison of the Internet penetration rates in Figure 6.2 clearly depicts a Digital Divide between the first ranked regions (North America, Australia and Europe) and the regions at the tail end (Asia and Africa).

By comparing the Internet penetration rate in developed and developing countries[1] between 1997 and 2007 (Figure 6.3), the Digital Divide seems to be an emerging problem. The difference in the penetration rate between developed and developing countries increased over this time, although the trend line shows, that the difference is growing slower and slower. This effect can be explained by comparing the historic growth rate of Internet usage in different world regions (Table 6.1). The growth rate between 2000 and 2002 in

[1]The categorisation in developed and developing countries in this figure is made by the UNESCO

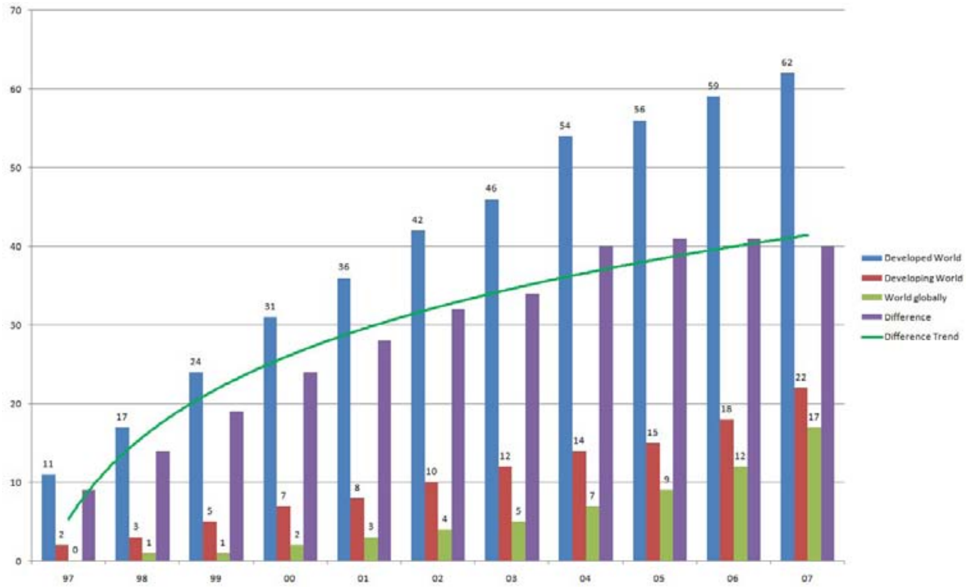

Figure 6.3: Internet penetration rate in developing and developed countries between 1997 and 2007 [11]

Table 6.1: Internet usage growth rate

World Regions	**00 - 02** [12]	**00 - 04** [10]	**00 - 05** [13]	**00 - 07** [14]	**00 - 09** [6]
Africa	52.1%	186.6%	423.9%	640.3%	1809.8%
North America	n/a	105.5%	110.3%	113.7%	140.1%
Latin America	n/a	209.5%	342.5%	466.2%	934.5%
Asia	62.3%	125.6%	218.7%	258.2%	568.8%
Europe	67.7%	124.0%	177.5%	203.6%	305.1%
Middle East	35.9%	227.8%	454.2%	491.1%	1675.1%
Oceania	71.5%	107.2%	134.6%	146.2%	177.0%
World Total	68.3%	125.2%	183.4%	214.0%	399.3%

Africa (52.1%), where most of the today's developing countries are residing, is significantly lower, than the growth rate in Europe (67.7%) or even Oceania (71.5%). But this ranking shifts over the years. Comparing the growth rate between 2000 and 2004 Africa's growth rate (186.6%) is already much greater than Europe's (124.0%) and North America's growth rate (105.5%). The most recent data (2009) show that the growth rate for all these countries with a low penetration rate (see Figure 6.2), especially Africa (1809.8%) and Middle East (1675.1%), is far higher than the growth rate of those countries which have already a high penetration rate, like for example North America (140.1%) or Oceania (177.0%). Therefore one can perhaps conclude that the Digital Divide is growing slower and slower, reaches its maximum, and scales down until it is almost vanished.

This would be very important for the disadvantaged countries. Fairlie states in his work that "the Digital Divide may have serious economic consequences for disadvantaged minority groups as information technology skills become increasingly important in the labor market and for education" [15]. In the future Internet access will be even more important, as Fairlie states: "Future economic, education, community participation and political advancement for these disadvantaged groups may depend on access to computers, the Internet and broadband technology" [15].

There are many reasons why the Digital Divide emerges. First of all the economic situation in a country plays an important role. But there are also other factors which influence the dimension of the gap between different countries. Some of these factors are explained in detail in the following sections.

6.2.3 Economic and other factors

At the end of the 20th century only few people could use respectively used the Internet, but since that time the access rate is growing rapidly. As shown in Figure 6.4 the number of Personal Computers has a similar progression rate as the usage of the Internet. Unfortunately, this burst of computer and Internet usage involves a huge difference among countries in the world. In contrast to the industrialized countries, where more then the half of the population possesses a PC and has access to the Internet, in most of the countries in Africa just a few people have access to the Internet and possess a PC. An other example which points out the gap is Russia. There, less than 14 citizens per 100 own a computer, whereas in the United States more than eighty out of hundred people own a computer [46].

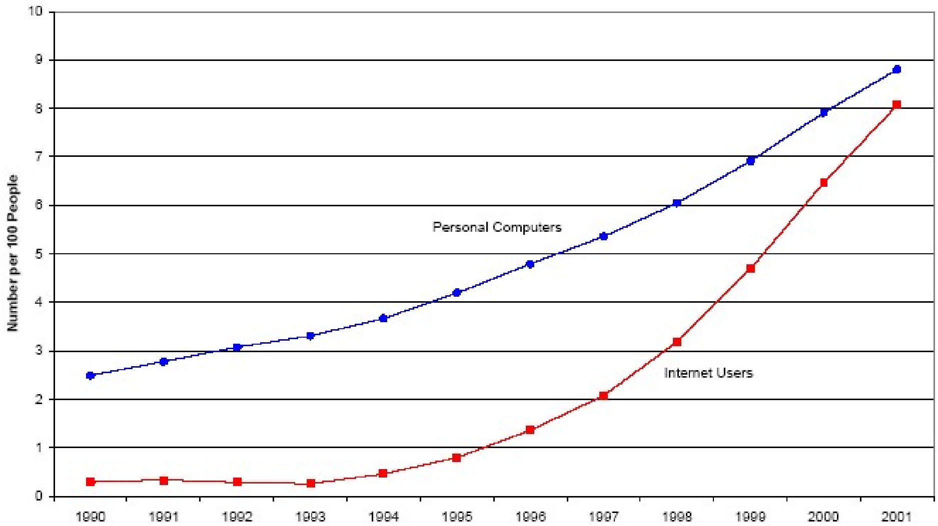

Figure 6.4: World computer and Internet penetration rates [47].

But what could be the factors, which are affecting this gap?

There are many different factors, which are affecting this unequal diffusion of technology in different countries. In this section we will point out the economic factor, which is income per capita, as well as other factors (years of schooling, regulatory quality and electricity consumption) which have been identified to be the main factors responsible for the Digital Divide [4]. Furthermore the importance of these factors regarding to the Digital Divide will be discussed. The International Telecommunication Union (ITU) and its World Telecommunication Indicator Database will help us analyzing the possession and the usage of Personal Computers and the Internet per 100 people [46]. To evaluate the influence of the other factors such as income per capita, telecommunication access and electricity consumption the statistics from the World Bank's World Development Indicators Database will be helpful [48].

6.2.3.1 Income per capita

Income per capita is the most important component of the Digital Divide, but as mentioned above it's not the only one. Perhaps, it is the key determinant of the computer possession. Some consumers would like to buy a computer because they see it's important, but they are not able to finance the purchase of one. The preference for having a computer also varies across individual characteristics, such as education level, age, urban/rural location, etc. Those factors aren't the only ones which influence the consumer's decision to buy a computer. Another important factor is the price of a computer. The examination of the hedonic price indices, which has been calculated for the ICT equipment in the US, provides evidence for large movements in the relative price of computing power.

The analysis of the computer and Internet penetration rates across countries shows interesting patterns, too. Table 6.2 reports countries with the highest and lowest computer penetration rates, followed by Gross National Income (GNI) per capita by 2006 and 2007. Canada has the highest computer penetration rate. The top nine also contains many European countries (Switzerland, Netherlands, Sweden, Norway and Denmark) and the United States. Australia and Korea (Rep. of) also have a high computer penetration rate. Internet penetration rates follow the same patterns across countries as computer rates. So the ranking of countries by Internet penetration rate is similar to the ranking of countries by computer penetration rate. The two exceptions here are Canada and Switzerland. Being the first respectively the second in the computer penetration rate ranking they are only on the 7th respectively 8th place in the Internet penetration rate ranking. Worth mentioning is that all of these countries have a high GNI per capita except of Korea. GNI per capita generally follows the same pattern across countries as computer and Internet penetration rates.

Hence, it is not surprising that income per capita is a powerful determinant of PC usage. Each $1000 increase in per capita income is associated with more than a one percentage point increase in the number of PCs per capita [4].

Table 6.2 shows estimations of computer penetration rate for the countries with the highest and in contrast with the lowest computer penetration rate in the world. The largest countries like Russia, Brazil, China and India have low computer and Internet penetration rates as well as low GNI per capita. In China, there are less than six Personal Computers per 100 people, less than seventeen Internet users per 100 people and the GNI per capita is less than 2500$. The computer and Internet penetration rate and GNI per capita are even lower in India. There are only 3.3 computers per 100 people, 7.2 Internet users per 100 people and the GNI per capita is only 950$. The bottom part of the table is mainly comprised of very poor countries, which most of them are in Africa. All of those countries have computer penetration rates of less than four users per 100 people. For example,

Table 6.2: Computer and Internet penetration rates for largest countries, for countries with highest and lowest penetration rates [46] and for countries with highest and lowest GNI per capita [48].

	Personal Computers (per 100 people)					Internet users (per 100 people)					GNI per capita, Atlas method (current US$)				
World Countries	#	2006	%	#	2007	#	2006	%	#	2007	#	2006	%	#	2007
United States	5	76.2	5.6	5	80.5	7	71	4.2	6	74	5	44'890	2.7	5	46'090
Sweden	4	83.6	5.4	4	88.1	1	86	-7.2	4	80	4	45'210	5.9	4	47'870
Denmark	6	69.6	-21.1	9	54.9	2	83	-2.9	3	81	3	52'560	5.5	3	55'450
Switzerland	2	86.5	6.1	2	91.8	6	71	1.6	8	72	2	58'790	3.5	2	60'820
Australia	7	60.7	0.0	7	60.7	9	51	4.1	9	53	8	33'190	7.7	8	35'760
Norway	8	59.4	5.9	6	62.9	3	81	5.1	1	85	1	68'270	13.3	1	77'370
Korea (Rep. of)	9	53.2	8.3	8	57.6	5	74	2.2	5	76	9	18'950	11.9	9	21'210
Canada	1	87.6	7.6	1	94.3	8	70	3.5	7	73	7	36'880	7.5	7	39'650
Netherlands	3	85.4	6.8	3	91.2	4	81	4.0	2	84	6	42'490	7.4	6	45'650
Mexico	9	13.6	5.9	10	14.4	10	20	6.4	9	21	11	8'740	7.6	11	9'400
Brazil	10	16.1	0.0	8	16.1	11	31	13.2	10	36	9	4'820	25.7	9	6'060
Russia	8	12.2	9.0	9	13.3	9	18	17.1	9	21	10	5'800	29.8	10	7'530
China	7	4.3	32.6	7	5.7	8	10.6	52.1	8	16.1	8	2'010	19.9	8	2'410
Indonesia	4	1.5	33.3	4	2	7	9.0	23.6	7	11.1	7	1'410	17.0	7	1'650
Nigeria	3	0.8	0.0	3	0.8	5	5.5	22.2	5	6.8	6	830	16.9	6	970
India	5	1.6	106.3	6	3.3	6	6.8	5.2	6	7.2	5	820	15.9	5	950
Bangladesh	6	2.2	0.0	5	2.2	2	0.3	9.5	1	0.3	2	450	6.7	2	480
Benin	2	0.4	75.0	2	0.7	4	1.5	16.2	4	1.8	4	570	7.0	4	610
Cambodia	1	0.3	0.0	1	0.3	3	0.5	4.2	3	0.5	3	500	10.0	3	550
Ethiopia	2	0.4	75.0	2	0.7	1	0.3	19.1	2	0.4	1	190	15.8	1	220

in 2007 the computer penetration rate in Canada was more than 314 times larger than the penetration rate in Cambodia. The three countries Benin, Cambodia and Ethiopia have also extremely low computer penetration rates. In Cambodia, there are only 0.3 computers per 100 people, and in Benin and Ethiopia there are only 0.7 computers per 100 people.

The bottom part of the table shows the countries with the lowest computer and Internet penetration rates. Astonishingly, there are more Internet users per 100 people than Personal Computers per 100 people. Like in the upper part of the table with the countries with the highest penetration rates, the place in the Personal Computers per 100 people ranking correlates with the place in the Internet users per 100 people ranking. In Indonesia for example the Internet penetration rate is six times higher than Personal Computers per 100 people but it is still very low (only 11.1 Internet users per 100 people). In Nigeria there are only 6.8 Internet users per 100 people but this is six times as high as the rate of Personal Computers per 100 people. Worth mentioning is that all of these countries have a low GNI per capita.

Most notably and a surprising result is that the ranking of GNI per capita of a country is roughly similar to the ranking of the same country by Personal Computers per 100 people or Internet users per 100 people. This result applies to all countries regardless of whether they have a high or a low GNI. For example, in 2007 Switzerland has the second place in the GNI ranking and the same place in the Personal Computer per 100 people ranking.

The positive relation between per capita income and computer/Internet penetration rates may be partly due to relaxing the budget constraint, changing preferences, or liquidity constraints [4].

6.2.3.2 Years of schooling

Years of schooling, which measure the human capital difference, are very important in contributing to the Global Digital Divide. The world average gap in computer penetration rates from 9.9% to 14.4% results from differences in the education system. Around the countries in the world there are many different average numbers of years of schooling. For example, the average number of years of schooling in Sub-Saharan is 3.7 years, in Europe and Central Asia it is 8.3 years. The United States have the highest average years of schooling which is 12.1 years [49], [50].

The main problem is, that the use of computers and Internet requires substantial levels of education, so countries with low level of education will limit their demand. The demand for computers increases significantly with higher levels of education [51]. Education does correlate positively with the degree of PC and Internet use. But the degree to which the difference in Internet rates depends upon this variable is surprisingly small. For most of the regions, the fact of lower education causes only about the half of the magnitude of the gap attributable to differences in the educational systems [4].

6.2.3.3 Regulatory quality

Regulatory quality is an important component of the Digital Divide. There are many reasons for the importance of regulatory quality. One of them is that regulations set the rate of taxes. This affects directly the firms in that country. If the taxes are too high, firms are forced to leave because they can not survive. Hence the demand for ICT equipment in that country declines.

Regulatory quality in the different regions appears to contribute to the Global Digital Divide. This differences in most of the regions create a gap in computer penetration rates about 10%. The gap of computer penetration in the Middle East and North Africa caused by differences in regulatory quality is nearly 15%. In Europe and Central Asia the regulatory quality is more similar to the United States and therefore this factor is only 4.7% of the gap. The substantially larger magnitude of contributions from regional differences in regulatory quality is the most notable difference between the results for the Internet penetration rate gaps and those for the computer penetration rate gaps. 11.7% of the gap between Europe/Central Asia and the United States and 18.2% to 32.0% of the gap between other regions and the United States result from differences in regulatory quality [4].

Nearly one-third of the Internet penetration rate gap might be closed if countries in the Middle East and North Africa had similar regulatory quality as the United States [4]. The regulatory quality affects the Internet use negatively [52].

6.2.3.4 Electricity consumption

Electricity consumption is also one of the important components of the Digital Divide. It has a high importance because without electricity the Personal Computer would not work. According to Table 6.3 there is no relation between per capita income, electricity usage and Personal Computer penetration rate with the exception, that all countries that use more than 7000 kWh per capita have also higher computer and Internet penetration rates. This means that electric power consumption also contributes to the Global Digital Divide.

Table 6.3: Electric power consumption (2005-2006) and total population (2005-2007) for largest countries and for countries with highest and lowest penetration rates [48].

	Electric power consumption (kWh per capitapeople)			Population, total				
World Countries	2005	%	2006	2005	%	2006	%	2007
United States	13'701	-0.9	13'582	295'561'000	0.9	298'363'000	1.0	301'290'000
Sweden	15'440	-1.4	15'231	9'024'040	0.6	9'080'505	0.7	9'148'092
Denmark	6'665	3.0	6'864	5'415'978	0.4	5'437'272	0.4	5'461'438
Switzerland	8'305	0.7	8'360	7'437'100	0.6	7'483'934	0.9	7'550'077
Australia	11'265	0.6	11'332	20'394'800	1.5	20'697'900	1.5	21'015'000
Norway	25'083	-3.1	24'296	4'623'300	0.8	4'660'677	1.0	4'709'153
Korea (Rep. of)	7'804	3.3	8'063	48'138'000	0.3	48'297'000	0.3	48'456'000
Canada	17'314	-3.2	16'753	32'312'000	1.0	32'649'000	1.0	32'976'000
Netherlands	6'988	1.0	7'055	16'319'850	0.2	16'346'101	0.2	16'381'137
Mexico	1'968	1.8	2'003	103'089'133	1.1	104'221'361	1.0	105'280'515
Brazil	2'016	2.8	2'072	186'074'634	1.1	188'158'438	1.0	190'119'995
Russia	5'785	5.8	6'122	143'150'000	-0.5	142'500'000	-0.3	142'100'000
China	1'783	14.5	2'041	1'303'720'000	0.6	1'311'020'000	0.6	1'318'309'724
Indonesia	509	4.0	530	220'558'000	1.1	223'041'632	1.2	225'630'065
Nigeria	127	-8.0	116	141'356'083	2.4	144'719'953	2.3	147'982'941
India	476	5.7	503	1'094'583'000	1.4	1'109'811'147	1.3	1'124'786'997
Bangladesh	136	7.9	146	153'122'039	1.5	155'463'091	1.5	157'752'512
Benin	75	-1.1	74	7'867'626	3.3	8'128'208	3.3	8'393'132
Cambodia	55	60.8	88	13'955'507	1.7	14'196'611	1.8	14'446'056
Ethiopia	34	12.0	38	74'660'901	2.6	76'627'697	2.6	78'646'128

The gap of computer penetration rate due to electricity consumption between Europe and Central Asia and the United States is 6.8% and the gaps between other regions and the United States are from 15.1% to 17.8%. Electricity is also essential for the Internet usage. The gap of the Internet penetration rate between Europe and Central Asia and the United States caused by regional differences in electric power consumption is 10.5% and the Internet penetrations rates between other regions and the United States are from 20.9% to 25.0% [4].

For these reasons, researchers and practitioners, as well as governments and organisation, have established many approaches and projects in order to try to reduce the digital gap. A small sample of them is presented in the following section.

6.3 Solutions

In this section we will present three approaches to bridge the Digital Divide. The OLPC project tried to bridge the gap by providing disadvantaged people with cheap laptops.

Another initiative used mobile phones to provide access to the worldwide communication to rural communities in Africa. The third approach uses Open Source Software to supply poor countries with cheap and flexible software.

6.3.1 OLPC

As mentioned in the previous sections, in a lot of developing countries, the public infrastructure like the power supply, the public road system or Internet access is not developed very widely. Therefore, even if the people in this countries could afford to buy a computer or a mobile phone, they are often not able to access the Internet because of the missing public access possibilities. Additionally, most of the people in these countries don't have the necessary skills to use ICT, because they were never taught how to use them. This problem starts in schooldays. In most developed countries it is a standard to use computers in school, even at a very young age. But in developing countries, most people don't get in touch with modern technological devices during their whole life.

In January 2005, at the World Economic Forum (WEF) in Davos, Switzerland, a former director of MIT's Media Lab, Nicholas Negroponte presented the idea of One Laptop Per Child (OLPC). The OLPC is a initiative that should reform the education of the schoolchildren in developing countries. The idea was to empower the children of developing countries by providing one laptop to every school-aged child. Additionally they were also provided with means for learning, self-expression and exploration, so that they should be able to learn by themselves [45]. In order that poor developing countries can afford to buy a huge amount of laptops, the goal of the OLPC project was to develop a new laptop that can be sold for only $100. Negroponte "estimated that up to 150 million of these laptops could be shipped annually by the end of 2004" [24].

In collaboration with major IT industry players, for example Google, AMD, or Red Hat, OLPC invested 20 million dollars and developed an extraordinary hardware innovation, the XO laptop [31]. This laptop reflects hardware innovation in the power supply, display, networking, keyboard and touchpad to provide a durable and interactive laptop [42]. As shown in Figure 6.5, some components of the XO laptop make it especially applicable for using them in developing countries [43]. These components are:

- **Power supply:** There are two choices of batteries available. According to OLPC, both of them hold their charge for at least four times longer than a standard laptop battery. In countries with electricity supply, the laptop can be used with a 18W power adapter. If a country does not have access to the power grid, there are many different ways to recharge the battery. For example, it is possible to recharge it with a solar panel, a hand crank (similar to those used on windup radios), a foot pedal, or a pull-string recharger, similar to a starter cord on a lawnmower. This pull-string recharger is very effective. Just one minute of pull-string gives a 10 minutes of charge. It is also possible to group recharging stations at schools or similar places where students can hook up their batteries to solar panels or car batteries.
- **Display:** The children can choose between a full-colored display that is similar to any other Liquid Crystal Display (LCD) and a black-and-white display that allows

Figure 6.5: XO features

them to read text also in direct sunlight. The screen can also be swilled around so that the laptop can be used as an ebook reader or a game console.

- **Networking:** Using standard wireless protocols, the laptops are able to form a mesh network which allows all connected laptop sharing Internet access, as soon as only one of them is connected to the Internet.

- **Keyboard:** The sealed green rubber keyboard is waterproofed and its size is designed for a child's hand. The keyboard can be customized for different languages and has a standard QWERTY layout.

- **Software:** The laptop has a bespoke Linux operating systems developed by the leading Open Source Software company Red Hat. Its user interface, known as Sugar, is particularly suitable for children and not very complicated.

- **Other components:** There are also other components that have made this laptop more useful and attractive for children. For example, a camera embedded in the display, allows children to video chat in mesh networks, but also to take pictures or use it as a light meter for school projects. The three USB data ports can be used to plug in keyboard and mouse or other devices. The SD memory card slot can be used to expand the memory capacity or to load new software. When the laptop is closed, the Wi-Fi antennas cover all data ports, preventing water and dust from getting inside.

In December 2007 the OLPC pilot phase started. A lot of high-level officials from developing countries were really enthusiastic about OLPC and committed to purchase a huge amount of XO laptops. Studies in six countries reported positive changes "such as increased enrollment in schools, decreased absenteeism, increased discipline and more participation in classrooms" [42]. The main problem was that these studies were not neither made systematically nor by independent researchers. Nevertheless, large purchase were made by Peru (640'000), Uruguay (420'000), Rwanda (120'000) and many other developing countries. At the same time OLPC launched a similar project, called Give One, Get One (G1G1). The goal of this project was to sell two XO laptops for $399 to a U.S. citizen, but send one of this two laptops to a child in a developing country. The first phase of this new project was also a huge success with 167'000 units sold [44].

Despite the announcement during the World Economic Forum 2005, OLPC was not able to produce the XO laptop for only $100. Actually, one unit costs $199 according to the OLPC website[2]. Because of this price increase some countries lost interest. Nigeria for example could not keep its promise to purchase 1'000'000 units, partly because of the increased price [32]. But there were also many other problems that reduced interest in OLPC. Insufficient distribution and support, only marginal training for teachers, short-handed sponsor development of local-language software and teachers' resistance to the new technology are only some of them [42].

Also the PC industry was not pleased with the XO laptop, because it was a potential thread to the PC industry in emerging markets. The reaction of the PC industry was very harsh. Craig Barrett, former CEO of Intel, called the XO a gadget, that no one really wants [33]. Nevertheless, in 2006, Intel introduced also a small laptop, called Classmate. This laptop was clearly designed for developing countries and can be purchased for $230 - $300 [42]. Intel was able to settle deals to sell "hundreds of thousands of Classmates in Libya, Nigeria, and Pakistan, some of the very countries OLPC was counting on" [42]. Microsoft also tried to get a market share and decided to sell Microsoft Windows, Microsoft Office and some educational software for only $3 per copy if used on computers in schools of developing countries. Because of this aggressive marketing strategy and because of the demands of some governments, OLPC was forced to allow Windows as a possible operating system on the XO laptop.

Even though the OLPC project could not fully meet its high expectations, it has prepared the market for low-cost laptops. Other manufacturers followed OLPC and Intel and are now producing their own low-cost laptops that are at least theoretically affordable for developing countries.

6.3.2 Mobile phones

The governments of developing countries play an important role in bridging the Digital Divide. It is particularly important that the governments raise funds to build a national infrastructure that will allow the population to access the Internet and other forms of ICT. Unfortunately, it is very expensive and time consuming to build fixed line telephone systems infrastructure and most of the developing countries don't have enough money for these investments. As a result a good deal of the people in developing countries don't have access to the Internet, are suffering from low bandwidth and have to pay high line installation fees.

In recent years, researchers have spotlighted mobile phones as an instrument to solve these problems and overcome the Digital Divide. Kauffman and Techatassanasoontorn

[2] http://www.laptop.org/en/participate/ways-to-give.shtml

explain the increasing interest in mobile phones: "Due to their affordability, popularity, and fast infrastructure implementation, digital wireless phones offer several benefits to developing countries [...]" [9]. Mobile phones have become a serious competitor to the poor maintained government-owned fixed telephone line systems. Particularly by developing and implementing the third-generation (3G) standards, mobile phones have become very interesting devices, because using these standards, it's possible to use high speed data services like web browsing, video streaming, file download or location-based services. These possibilities render them interesting for the users. The standardization process of 3G started in 1998 and was lead by the ITU [16].The world's first commercial launch of a 3G network took place in Japan on 1 October 2001 [17].

For these reasons mobile phones are widely used in developing countries, as an UNCTAD[3] report mentioned in 2008: "Among the forms of ICT, mobile phones are most widely used in developing countries. Mobile phone subscribers have almost tripled in such nations over the last five years, and now make up some 58% of mobile subscribers. Developing countries in Asia and Africa had the highest growth rates in mobile phone penetration worldwide" [18]. These statements can be confirmed by looking at the annual growth rate statistics in mobile subscribers by the ITU[4] (Figure 6.6). One can see that Africa's growth rate is far above the growing rates of all other continents, followed by Asia whose growing rate is above the world average too. It is also important to mention that Africa has a very high amount of mobile subscribers, compared to the total sum of telephone subscribers. This fact also confirms the assumption that mobile phones are particulary widespread in developing countries [9].

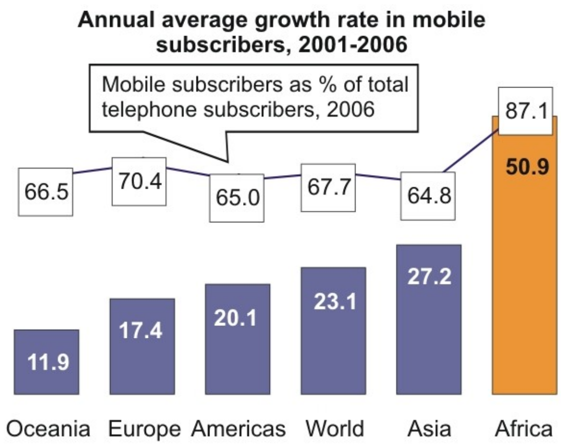

Figure 6.6: Annual average growth rate in mobile subscribers, 2001 - 2007 [19]

This wide proliferation of mobile phones in developing countries has arisen interest of the mobile phones industry. Motorola, one of the biggest handset manufacturer worldwide, believes that "the mobile device will close the Digital Divide in a way the PC never could" [20]. So, the mobile phone industry has launched a lot of projects in order to distribute affordable mobile phones in developing countries. "The aim is to show that a 'win-win-win' scenario is possible, in which users get affordability, manufacturers and operators sell more

[3]United Nations Conference on Trade and Development
[4]International Telecommunication Union

handsets and airtime, and governments raise their tax revenues", as Robert Conway, CEO of the GSM Association states [21]. The remainder of this section will shortly present an exemplary project launched by the GSMA (GSM Association).

The GSMA is an association of over 800 GSM mobile phone operators, manufacturers and related companies from across the world. It focuses on standardisation and promotion of the GSM mobile telephone system [22]. The GSMA has launched an initiative called 'Bridging the Digital Divide' which consists of several projects that "seek how the high speed GSM data network (GPRS, EDGE, or 3GSM) can be used to bridge the Digital Divide. The focus is on innovative solutions that provide not only a channel to the Internet but a suite of relevant services in a cost efficient, sustainable and scalable manner" [23].

One of these projects took place in Algeria, where GSMA supported a local provider, called Djezzy, in establishing public telephones in a rural region of Algeria. In this region there are living both settled and vagrant people with very low income. So, with the help of GSMA, Djezzy made "subsidised Psitek[5] units available to local communities, either by placing them in small villages, or at the network of livestock and irrigation water points situated throughout the area" [23]. Afterwards, site managers were chosen which are responsible to operate the phones at subsidised rates. According to the GSMA Development Fund [23], this project had some strong impacts. "All parties improve their social and economic development capacity, and help to make a real difference to the lives of people in Algeria's rural communities". The people in the local communities don't have to travel a long way to make a call anymore, the local site managers are provided with a stable revenue stream and the local mobile operator Djezzy can benefit from its first mover advantage.

6.3.3 Open Source Software

Another approach to bridge the Digital Divide might be Open Source Software (OSS). This term is being used in a confusing way nowadays, therefore it's essential to explain the proper meaning of OSS. First of all, OSS has to be distinguished from Freeware, which is as well free to use but without the distribution of the source code (that means no modifications are possible). According to the Open Source Initiative website[6], OSS must satisfy the Open Source Definition (OSD) with its 10 criteria [35]. Four of them are essentially important concerning the Digital Divide:

- **Free distribution:** The license shall not restrict any party from selling or giving away the software and shall not require a royalty or other fee for such sale.
- **Source code:** The program should be distributed in compiled form as well as in source code. Obfuscated source code ist not allowed since it makes a modification by the programmer impossible. This is an important point which allows programmers to change the source code to be up to the mark.

[5]Psitek is one of Africa's biggest manufacturer of telecommunication solutions (`http://www.psitek.com`)

[6]Open Source Initiative (`http://www.opensource.org`)

- **Derived works:** Derived works as well as modifications of the program should be allowed.
- **No discrimination against fields of endeavor:** Everyone should be allowed to use the program in any specific field of endeavor (like business, genetic research, etc.).

As seen above, there are a lot of advantages of using OSS. According to Richter, there are four benefits of OSS especially for developing countries [36]:

- **Transfer of knowledge:** Since the source code is available, programmer in disadvantaged countries can learn from the code locally. This means that learners can become experts.
- **Marginal costs:** Proprietary operating systems are expensive. In African countries like Rwanda for example, a WinXP license costs more than 31 average monthly salaries [37]. The modification, if necessary, of OSS furthermore generates new jobs and is not so expensive because of the low salaries in developing countries.
- **Technological independency:** Upgrades, bugfixes and support create a dependency of the buyer on the software manufacturer which lasts many years. In addition to that there is always a vendor lock-in when using software, so they can't change to other operation systems later easily. In contrast to that Open Source Formats (OSF) are all well documented and based on standards. Furthermore OSS can be supported locally. Another argument for OSS is, that Linux for example doesn't need the newest (and expensive) high-performance hardware like the newest Windows does.
- **Software localization:** Proprietary software is only available in languages which are profitable for the manufacturer. Since the source code isn't distributed, a translation is possible only by the manufacturer. This is different with OSS, which can be localized by a user (programmer) due to the available source code.

There are examples where organizations and even governments have already tried to bridge the Digital Divide with OSS. In South Africa, The Zuza Software Foundation has tried to overcome the localization problem with proprietary software. South Africa has 11 official languages, but most computer software are only available in English. The second language, Africaans, is poorly, the other nine are not supported at all. The aim of The Zuza Software Foundation was to establish a project to translate software into various official languages of South Africa [38]. Due to the open source translations are possible at all. The opensource software translation project[7] was born. Mozilla Firefox, Mozilla Thunderbird and OpenOffice are now available in several South African official languages.

On the governmental side there's one country which is already using OSS in the e-government area on a great extent. To bridge the Digital Divide Brazil began to use OSS in education, cybercafes and in the public sector. The government made available

[7] Opensource software translation project (`http://www.translate.org.za`)

workstations based on Linux in cybercafes in different areas. In Sao Paulo City for example, every citizen can use a computer for one hour a day to surf the Internet or to do homework [39]. The users hope that the new skills will help them during the search for a job. Since the unemployment rate is very high, getting a job is the number one worry there. In 2007, the Brazilian Public Software Portal[8] was created. 34 software solutions in different areas such as education, health and public administration are available on this portal [40].

It is obvious that OSS can't bridge the Digital Divide alone. If you can't afford a computer or the Internet costs you don't care about software. Mitch Kapor, chair of the Open Source Application Foundation and developer of Lotus 1-2-3, notices that OSS is not the stand alone solution but it helps narrowing the gap. The remaining barriers are access to a computer and access to broadband [41].

6.4 Conclusion

In the following two sections we will shortly present the drawbacks of the approaches mentioned in Section 6.3. We will analyze which factors were responsible so that the projects weren't really successful. Afterwards, in Section 6.4.2, we will list additional barriers which will have to be beared down in order to bridge the Digital Divide.

6.4.1 Why didn't it work? What could have been done better?

As already mentioned in previous sections, the Digital Divide has some severe negative effects. Lack of access to ICT services and devices give rise to serious social and economic disadvantages. Countries which don't take advantage of the digital technologies can't really compete in the new digitalized global market. To make it worse, this problem will aggravate as the technological development progresses faster and faster. And there is no doubt that the technological progress will continue in the next decades. Referencing Moore's law [25] of doubling number of transistors on an integrated circuit every two years, Intel's CEO Craig Barrett thinks that this will last for the next few decades [26]. Success stories like for example South Korea [27] convince other developing countries that controversial actions in trying to bridge the Digital Divide will pay off and will urge on them to take similar actions.

But one has to be careful in taking the correct actions. It is more or less universally accepted that providing access to the Internet and other digital technologies is not the complete solution to bridge the Digital Divide. Nevertheless a lot of countries which become scared of falling behind in the digitalized economic system spend billions of dollars in massive infrastructure projects in order to provide access to the Internet. Warschauer [5] mentioned in his paper that many projects around the world which try to bridge the Digital Divide fail, because they "too often focus on providing hardware and software and pay insufficient attention to the human and social system that must also change

[8]Brazilian Public Software Protal (`http://www.softwarepublico.gov.br`)

for technology to make a difference". This is one of the reasons why the OLPC project mentioned in Section 6.3.1 could not meet its high expectations. The OLPC project developed a very innovative device but lacked in understanding the needs of the developing countries. So they developed a cheap device which no one really could use because "information technologies are not standalone innovations but system innovations, the value of which depends largely on an ecosystem that includes hardware, applications, peripherals, network infrastructure and services (such as installation, training, repair, and technical support). Deployment involves training teachers, creating software and digital content, delivering maintenance and support, and sustaining a long-term commitment. Such capabilities are in short supply in developing countries, and OLPC simply never had the resources to provide them" [24]. The same problem could cause the mobile phone project (described in Section 6.3.2) to fail. There the GSM Association provides a rural area in Algeria with shared phone access without spending a lot of effort in training and educating the people who are meant to use, operate and maintain these phones.

Open Source Software offers some great opportunities to disadvantaged developing countries. They can purchase it free of charge and it is possible to adapt it very easily to local standards, as South Africa did it. But nevertheless, Open Source Software can't be the only action taken to bridge the Digital Divide. It is necessary to establish some accompanying measure. In the first place, developers are needed which can modify and adapt the Open Source Software. For this purpose it is important to educate people and to teach them how to work with the software. Additionally it is necessary to guide the prospective users so that they are able to use the software.

All of the three approaches mentioned in Section 6.3 focus on one specific part of the Digital Divide problem. But in order to really bridge the gap, it is necessary to consider the whole situation. In the next section we will present some further barriers which will have to be beared down in order to improve the status quo.

6.4.2 Is the problem solvable at all?

In the previous section we presented why the projects described in Section 6.3 were not a huge success. All these projects concentrated almost exclusively on providing access to ICT devices and services without proper education and without understanding the specific situation in the developing countries. In order to launch a initiative with higher chances of success it is necessary to consider also other factors than only the access problem. Taglang discovers in his report at least 4 other barriers (in addition to insufficient infrastructure and education) which have to be beared down in order to bridge the gap [28]:

- **Lack of local information:** Most people using ICT want to handle information about their local community. In developed countries there is a lot of information-rich content available for almost all local communities, because of the high penetration rate, even in rural areas. But in developing countries, there is almost no local information available.
- **Literacy barriers:** In order to use ICT it is in the majority of cases necessary that the user can read at an at least average literacy level. According to the UNESCO

[29] the developing countries have a literacy rate of 76.4% in 2004. In Sub-Saharan Africa (59.7%) and South and West Asia (58.6%) almost every other citizen won't be able to use ICT properly in the first place.

- **Language barriers:** According to the newest statistics by Miniwatts [30], published in 2009, almost half of the users using the Internet are speaking English or Chinese. Most information in the Internet is made for these users and is therefore written in English. People who don't speak English are often not able to benefit from the advantages the Internet offers.
- **Lack of cultural diversity:** Studies have shown that cultural information is a major incentive for people to use the Internet [28]. But there is a lack of cultural diversity in the Internet. Only in countries with a high Internet penetration rate, it is possible, that every local culture has its own content in the Internet. In developing countries there are just too few peoples online to create this content. This lack of cultural information makes the Internet uninteresting for a lot of people.

Other researchers, as for example Compaine, believe that the Digital Divide is only a perceived gap. He argues that technological differences between countries are relatively short-dated and therefore they will dissolve even without any counteractive measures. According to Compaine, "at least two factors account for the rapid diffusion of Internet technology: steadily decreasing costs of use, and steadily increasing ease of use" [34]. Following this argumentation, the best approach to solve the problem of the Digital Divide is to do nothing at all and wait until the problem is solved by itself.

A lot of people argue that the Digital Divide is merely an economic problem. They claim that as soon as poor people living in developing countries can afford to buy a computer or a mobile phone and as soon as the governments in these countries are able to build the necessary infrastructure, the problem of the Global Digital Divide will be solved very fast. This argumentation ignores the fact that even if people can afford buying a computer or a mobile phone, they are not automatically capable of using it.

In summary it can be said that recent statistics show that the Digital Divide is an emerging problem, especially for developing countries, which won't be solved by itself. It is necessary that governments, societies, academia and enterprises from all over the world get involved with these problems and try to bridge the gap by establishing sustainable projects and approaches which account for the cultural diversity, the environment they are situated in and the people which should work with these projects.

Bibliography

[1] United Nations Conference on Trade and Development: E-Commerce and Development Report 2004; New York/Geneva, 2004.

[2] S. Cisler: Hot Button: Online Haves vs. Have-Nots, San Jose Mercury News, 01-16-2000, http://www.athenaalliance.org/rpapers/cisler.html, April, 2010.

[3] W. Chen, B. Wellman: The Global Digital Divide - Within and Between Countries; IT&Society, Vol. 1, Issue 7, Spring/Summer 2004, pp. 39-45.

[4] M. D. Chinn, R. W. Fairlie: The determinants ot the global Digital Divide: a cross-country analysis of computer and Internet penetration; Oxford Economic Papers Vol. 59, 2007, pp. 16-44.

[5] M. Warschauer: Reconceptualizing the Digital Divide, First Monday, Volume 7, Number 7, 2002.

[6] Internet World Stats: World Internet Usage Statistics News and Population Stats; http://www.Internetworldstats.com/stats.htm, March, 2010.

[7] Röller and Waverman: Telecommunications Infrastructure and Economic Development: A Simultaneous Approach; Discussion Paper FS IV 96 - 16, Wissenschaftszentrum Berlin, 1996.

[8] Dewan and Kraemer: Information Technology and Productivity: Evidence from Country-Level Data; Management Science Vol.46, No.4, April 2000, pp. 548-562.

[9] Kauffman and Techatassanasoontorn: Is There a Global Digital Divide for Digital Wireless Phone Technologies?; Journal of the Association for Information Systems Vol. 6, No. 12, December 2005, pp. 338-382.

[10] Internet World Stats accessed via Internet Archive: Wayback Machine; http://web.archive.org/web/20041010084212/www.Internetworldstats.com/stats.htm, April, 2010.

[11] Own illustration based on the graphic from International Telecommunication Union (ITU); http://www.itu.int/ITU-D/ict/statistics/ict/graphs/Internet.jpg, March, 2010.

[12] Internet World Stats accessed via Internet Archive: Wayback Machine; http://web.archive.org/web/20030606022021/http://Internetworldstats.com/stats.htm, April, 2010.

[13] Internet World Stats accessed via Internet Archive: Wayback Machine; `http://web.archive.org/web/20060613025315/http://www.Internetworldstats.com/stats.htm`, April, 2010.

[14] Internet World Stats accessed via Internet Archive: Wayback Machine; `http://web.archive.org/web/20070617080319/http://www.Internetworldstats.com/stats.htm`, April, 2010.

[15] Fairlie: Are We Really a Nation Online? Ethnic and Racial Disparities in Access to Technology and Their Consequences; Report for the Leadership Conference on Civil Rights Education Fund, National Poverty Center Michigan, 2005.

[16] Smith and Collins: 3G Wireless Networks; MaGraw-Hill, New York, 2002.

[17] HKTDC: World's first 3G launch on 1st October severely restricted; `http://www.hktdc.com/info/mi/a/imn/en/1X00GV87/1/International-Market-News/World-s-first-3G-launch-on-1st-October-severely-restricted.htm`, April, 2010.

[18] UNCTAD: Science and technology for development: the new paradigm of ICT; Information Economy Report 2007 - 2008, United Nations, 2007.

[19] ITU: Market Information and Statistics (STAT); `http://www.itu.int/ITU-D/ict/statistics/ict/index.html`, April, 2010.

[20] USA Today: Motorola, Nokia expect payoff in bridging the digitial divide; USA Today, January, 16., 2007, `http://www.usatoday.com/money/industries/technology/maney/2007-01-16-digital-divide_x.htm`, April, 2010.

[21] Robert Conway: White Paper; Principal Voices, `http://www.principalvoices.com/voices/robert-conway-white-paper.html`, April, 2010.

[22] GSM World: About us; `http://www.gsmworld.com/about-us/index.htm`, April, 2010.

[23] GSMA Development Fund: Bridging the Digital Divide; `http://gsmworld.com/documents/Voice_Replication_Guide.pdf`, April, 2010.

[24] Kraemer, Dedrick and Sharma: One laptop per child: vision vs. reality; Communications of the ACM, Vol. 52, No. 6, 2009.

[25] Moore: Cramming more components onto integrated circuits; Electronics, Volume 38, Number 8, April 19, 1965.

[26] Kanellos: New life for Moore's Law; cnet, `http://news.cnet.com/New-life-for-Moores-Law/2009-1006_3-5672485.html`, April, 2010.

[27] Woyke: Korea Bridges Digital Divide; Forbes, March 4, 2009, `http://www.forbes.com/2009/04/02/Internet-broadband-korea-technology-korea-09-broadband.html`, April, 2010.

[28] Taglang: Content and the Digital Divide: What Do People Want?; Benton Foundation, http://www.technologyforhumanity.org/articles/contentanddd.html, April, 2010.

[29] UNESCO: Literacy for life; EFA Global Monitoring Report 2006, Paris, 2005.

[30] Internet World Stats: Top Ten Internet Languages; http://Internetworldstats.com/stats7.htm, April, 2010.

[31] BBC News: Intel and $100 laptop join forces; July, 2007, http://news.bbc.co.uk/2/hi/6897950.stm, April, 2010.

[32] Stecklow and Bandler: A little laptop with big ambitions; Wall Street Journal, November 24, 2007, http://online.wsj.com/public/article/SB119586754115002717.html, April, 2010.

[33] Reuters: Intel calls MIT's $100 laptop a 'gadget', CNET News, December 9, 2005, http://news.cnet.com/Intel-calls-MITs-100-laptop-a-gadge/2100-1005_3-5989067.html?tag=html.alert, April, 2010.

[34] B. Compaine: The Digital Divide: Facing a Crisis or Creating a Myth?; MIT Press, Massachuesetts, 2001.

[35] Open Source Initiative: The Open Source Definition (Annotated); http://www.opensource.org/docs/definition.php, April, 2010.

[36] M. Richter: Fair Code - Freie/Open-Source-Software und der Digital Divide; Open Source Jahrbuch 2006, www.opensourcejahrbuch.de/download/jb2006, April, 2010, pp. 371-380.

[37] R. A. Ghosh: Closed vs. Open Software; First Monday, Volume 8, Number 12, 2003.

[38] L. Martindale: Bridging the Digital Divide in South Africa; http://www.linuxjournal.com/article/5966, April, 2010.

[39] BBC News: Brazil bets on Linux cybercafes; December, 2003, http://news.bbc.co.uk/2/hi/technology/3250876.stm, April, 2010.

[40] GHDonline: Brazilian Government experience with Open Source; December, 2009, http://www.ghdonline.org/tech/discussion/brazilian-government-experience-with-open-source-d/, April, 2010.

[41] MuniWireless: Mitch Kapor sees open source as bridge to digital inclusion; July, 2007, http://www.muniwireless.com/2007/07/22/mitch-kapor-sees-open-source-as-bridge-to-digital-inclusion/, April, 2010.

[42] Kenneth L. Kraemer, Jason Dedrick, and Prakul Sharma: One Laptop Per Child: Vision vs. Reality; Communications of the ACM Vol. 52, No. 6, June 2009, pp. 66-73.

[43] BBC NEWS: Factfile: XO laptop; July, 2007, http://news.bbc.co.uk/2/hi/technology/6679431.stm, July 2007.

[44] G1G2: OLPC Give 1 Get 1 Program: XO-1 Laptop USA Sales; September, 2007, www.olpcnews.com/countries/usa/olpc_xo_laptop_sale.html, April, 2010.

[45] OLPC: OLPC - One Laptop per Child - Pictures, Videos and News;www.olpc.com, April, 2010.

[46] International Telecommunications Union: World Telecomunication Indicators Database, 2005-2007.

[47] International Telecommunications Union: World Telecomunication Indicators Database, 1990-2001.

[48] World Bank: World Development Indicators, World Bank, Washington, DC, 2005-2007.

[49] Caselli, F. and Coleman, W.J., II: Cross-country technology diffusion: the case of computers; American Economic Review, 91, 328-35, 2001.

[50] Pohjola, M.: The adoption and diffusion of ICT across countries: Patterns and determinants; The New Economy Handbook; Academic Press, New York, 2003.

[51] Fairlie, R.W: Race and the Digital Divide; Contributions to Economic Analysis & Policy; The Berkeley Electronic Journals, 3, Article 15, 1-38, 2004.

[52] Wallsten, S.: Regulation and Internet use in developing countries,; EconomicDevelopment and Cultural Change, 53, 501-23, 2005.

[53] Internet World Stats: Asia Internet Usages Stats and Population Statistics; http://www.internetworldstats.com/stats3.html, Mai 2010.

Chapter 7

Economic Challenges and New Opportunities using Cloud Computing

Stefan Badertscher, Frank Neugebauer, Daniel Meier

In this work we discuss today's reception of Cloud Computing and offer a differentiated view through the use of three use cases. We establish a working understanding of Cloud architectures and service types before analyzing popular Cloud Computing Services and their key points. Subsequently we present three fictional use cases – e-Mail in a large company, a growing book shop and scientific computation – and show which economic aspects arise and under which circumstances Cloud services are cheaper than regular best practice solutions. We identify some economic issues of Cloud Computing and try to offer some solutions to these problems.

Contents

7.1 Introduction and Motivation

What is Cloud Computing? When hearing this term the first time, most people usually just think of just another buzzword in the huge IT landscape. After further consolidation readings into the area of Cloud Computing, it rapidly becomes clear that this term stands not only just for something new, even more for something that is said to have revolutionary power. The following quote is from Achim Berg, the CEO of Microsoft Germany, about Cloud Computing at the Cebit 2010 in Hannover:

"Cloud Computing has the chance to change the world as similar as the Personal Computer did many years before." [17]

There are also critical voices which consider Cloud Computing just as some kind of a new marketing hype, which lead Richard Stallman to the following statement [2]:

"It's stupidity. It's worse than stupidity: it's a marketing hype campaign. Somebody is saying this is inevitable – and whenever you hear somebody saying that, it's very likely to be a set of businesses campaigning to make it true. "

Cloud computing is not only very interesting in the fields of technology it encompasses; its aspects from an economic point of view are also quite interesting. General expectations are that Cloud Computing should be able to reduce the IT costs in large companies by 30% [17]. Nowadays – since the cost pressure on the global market has strongly increased – such a cost reduction through Cloud Computing could mark the key enabler in the way of establishing this new kind of technology. As mentioned, it seems that Cloud Computing is supposed to be something revolutionary which is worth further investigation. In regard to the development of the Internet, Cloud Computing seems to be the next step in the internet evolution [16]. A strong motivation to choose this topic was our interest in the future development of the whole IT landscape. Today companies need to spend a big share of the budget for new IT solutions every year. The high cost pressure among the global competition will force companies towards the implementation of new and revolutionary technologies if they want to stay competitive. Therefore we believe that Cloud Computing could be a possible approach as well as a very interesting technology, which made our motivation for this topic even stronger. Based on this statement, it turns out that this new technology is already quite popular in the wild, as we will discuss later in this paper. The focus of this paper is not just based on academic work; it also includes practical case studies and reports. Since Cloud Computing is a straight emerging phenomenon in practice, those aspects should not be ignored. With the combination of academic and practical work a more accurate and detailed look at the problem can be accomplished.

This paper is structured in the following way: The first part covers an overview over the large area of Cloud Computing, where the basics principles are introduced. To achieve a deeper understanding of the topic and in order to make educated interpretations of economic challenges and fields of opportunities, there will be as much Cloud Computing theory as needed for the following sections. Hereby the focus will be set on different aspects of Cloud Computing. The whole paper will be rounded up by several interpretations of the covered aspects and by future prospects.

7.2 Cloud Computing – A brief Overview

The questions about what Cloud Computing is cannot be answered in just a couple of sentences, because it is told being some kind of paradigm shift [16]. A figurative start into the topic would be to consider the following picture in order to create a first abstract understanding what Cloud Computing is. Basically, it means that all the relevant software and data is managed on a central server which refers to a specific task [2]. Hereby the cloud stands for the Internet [25].

Figure 7.1: First Overview of Cloud Computing [7]

In other words, a user can access his data at (almost) every time using multiple devices as the following picture shows. Often large enterprises need these following opportunities for multi-accessing data. This practice is for example used by force.com [21].

Figure 7.2: Access Model [7]

Both pictures visualize the trend that data and software will not be stored locally anymore. In fact the aim of Cloud Computing is to store data and software somewhere in the cloud, which means in the Internet. Such a cloud can be located thousands of miles away from the current user [26].

Later the composition of Cloud Computing and its different levels will be explained.

7.3 Aspects of Cloud Computing

The brief overview has roughly drafted what Cloud Computing is about. Now the basics are introduced including a closer look at more detailed aspects. After describing and precising the definitions and classifications of Cloud Computing, the paper will focus on new opportunities and the economic challenges that go together with Cloud Computing. Based on the composition of this paper, some aspects maybe may already be mentioned before being explicitly covered. Anyhow the detail coverage will be placed in the corresponding sections.

7.3.1 Architecture and Prinicples of Cloud Computing

As seen so far, Cloud Computing services are offered in the clouds [26]. This leaves the question when the exact term “Cloud Computing” is appropriate. Is it only appropriate for a specific kind of service or does it even encompass other available forms of Cloud Computing like mixed forms with traditional services? In fact Cloud Computing is quite difficult to define exactly since there are many dimensions in which they can be separated; luckily there are two very common perspectives for the characterization [22], [3]. Interestingly there is a different name designated for the two most common categories. It can be based on the characterization from an organizational or a technical view [3], as well as a characterization on Service Boundary and Service Type [22]. But there is no need to be confused, as these are just different names for the same concepts, as we will see in the closer characterization. In the following we will have a closer look on these two different approaches, since they are quite essential for a deeper understanding in Cloud Computing.

7.3.2 Service Boundary

Taking a closer look at the classification of Cloud Computing using the Service Boundary approach, three substantial types [22] can be identified. Hereby the organizational entities are divided in users and providers. The technical behavior is not the key principle in this kind of classification. In this classification, the behavior of the service with providers and users represents the key fact. The classification is subdivided into the following schema.

The meaning of this picture will become clear after having considered the following description.

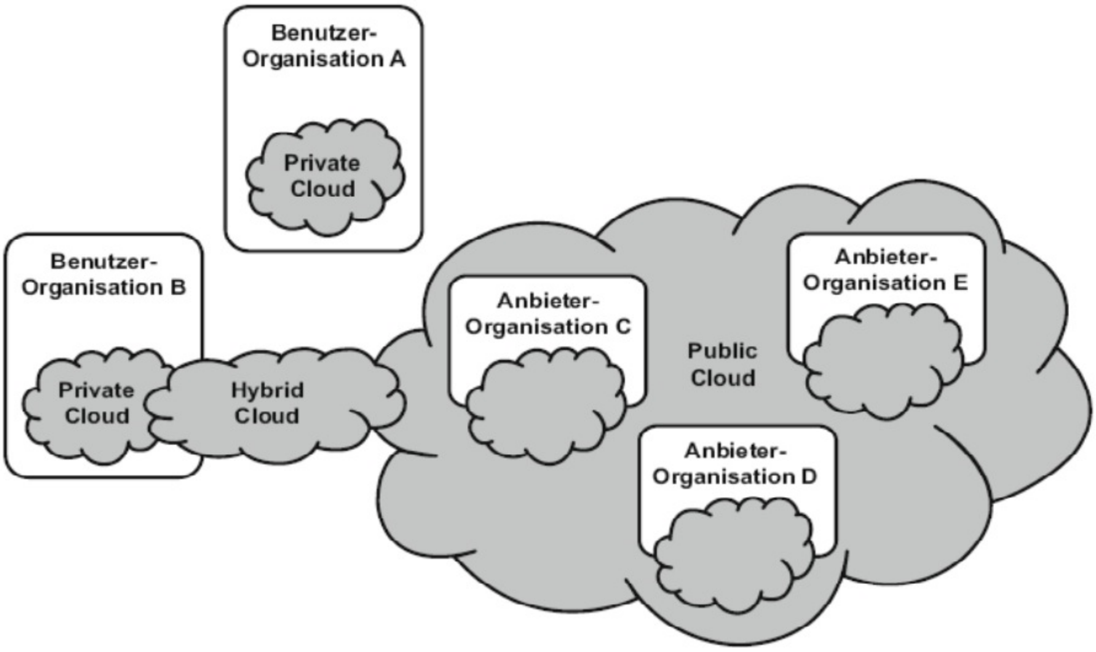

Figure 7.3: Overview of Public, Hybrid and Private Cloud Classification [3]

7.3.2.1 Public Cloud

Public Clouds can also be called "External Clouds" [3]. In such a case, the service is offered for external purposes and parties [22]. Roughly distinguished, if providers and users of cloud services do not reside in the same organization, it is called a Public Cloud. This generally means that such a cloud service is open for the public [3]. For example it can be offered over a web-interface, where the user can choose and specify his needs [3]. Often there are no fixed costs in the pricing model, it is considered as the state of the art that an entity using a service from a Public Cloud only has to pay for the resources which were actually consumed [3].

7.3.2.2 Hybrid Cloud

In Hybrid Clouds, resource sharing and the usage of services between public and private clouds over a secured network are typical [22], [3]. Usually several functionalities are constantly outsourced from the private to the public cloud. Especially when the need for extra resources is very high, an outsourcing of services to the public cloud happens quite often. Some examples of Hybrid Clouds would be using a Virtual Private Cloud (VPC). Such services are operated by Google and Amazon [22]. Due to security considerations, business critical data and programs should not been stored outside of the organizational boundary and therefore they should not be outsourced to the public cloud [3].

7.3.2.3 Private Cloud

It can easily be guessed that Private Clouds are the opposite of Public Clouds, where the provider and the user have to be in the same organization [3]. Such services are built and operated within and only within the own organization [22]. Other common terms for Private Cloud are *Internal Clouds* or *IntraClouds* [3]. When operating a Private Cloud, the using entity does not need to transfer its critical (business) data outside of its own organizational boundaries, which is a great advantage in contrast to Public Clouds. However this advantage has to be paid with the use of own resources, which can be very expensive [3]. It is obvious that the same interfaces in public and private clouds are needed in order to upscale own cloud software in the landscape of Public Clouds, if this is desired [3].

7.3.3 Service Type

In order to define Cloud Computing in the view of Service Type [22], there are the following three key facts to be mentioned, which are covered in more specific detail [11]. Following functional features, Cloud Computing can be separated into the following three different service levels [20], [3]. Hereby the characterization is made from a technical view and the paradigm is called *XaaS*, which means everything-as-a-Service [3]. (Obviously the X stands for any undefined service, also for everything.) After the reading, it will turn out that the following three items are ordered in a top down approach. First *SaaS* as the most abstract part situated on the top level, then *PaaS* in the middle since *SaaS* bases on it. The *PaaS* itself bases on the *IaaS*. Last but not least, the *IaaS* encloses all the hardware devices. The following more detailed description of such service types are the fundamental concept in understanding Cloud Computing.

- **Software-as-a-Service (SaaS):** Software-as-a-Service is the most common service in Cloud Computing. By now Payroll, Supply Chain Management and Customer Relationship Management are often performed over the Internet. These Services can be directly forwarded to the user web-interface, since the infrastructure components and the platform, which is below the specific service, does not need to be managed. About 45% of organizations are beginning to implement or increasing the use of Software-as-a-Service [11]. Such applications, for example Oracle CRM On Demand, Netsuite and Salesforce.com are hosted in the data center of the SaaS provider and are normally accessible over the internet [20]. These top level services may run on other PaaS and / or IaaS providers, using again their infrastructure as a service [20]. Managing centralized software provides advantages for the provider and as well for the user. Due to centralization, it is a lot easier to administer versioning and updating the used software [2]. Also the occurring problems when using sophisticated software can be reduced by the user. As with the SaaS the need for powerful hardware is quite low, such software can be operated by using a Thin-Client Architecture in a company [5].

- **Platform-as-a-Service (PaaS):** Platform-as-a-Service are platforms which can be used over the web by several developers in order to develop new native Cloud Applications, therefore it is an application development and deployment platform. It uses the underlying infrastructure in an abstract level, so that a developer does not have to bother with performance topics or old hardware components [20]. Some middleware components and tools for the application development process are situated in the PaaS. It is said, that this market will grow to a volume of 15 billion US$ in about eight years, because the companies' interest in running its own middleware will become low [11]. Often the PaaS have an own API or a specific programming language, for example the Google AppEngine is a PaaS, where developers have to use Java or Python in order to write applications [20]. The underlying infrastructure is kept up to date by numerous upgrades [21].

- **Infrastructure-as-a-Service (IaaS):** The Infrastructure-as-a-Service contains all the needed hardware, in other words it is the very basis on which the whole construct is founded and it subsumes all the servers, networks and storage devices [20]. It further allows to scale the capacity of the infrastructure accordingly to the requested tasks [11]. This means that a simple task like writing an e-Mail will request much less capacity from the IaaS Provider than for example a video rendering program. Based on consumption billing [20], the load balancing can be done in calibration with flexible price structures [20]. Nowadays about 25% of companies are using an own IaaS or request it from an IaaS provider [11]. The IaaS provider focuses only on keeping his infrastructure running. Managing the data stored on these hardware devices is not his task, as well as updating software, which must be done by the service user and not by the service operator [20]. For example, when the local hard disk drive of an service user crashes, none of his data will be lost as a remote copy is stored in the cloud. But often a hard drive is not needed and the operations take place on the data of the IaaS providers data center [26]. Some examples of IaaS providers are Amazon Web Services Elastic Compute Cloud (EC2) and Secure Storage Service (S3).

- **Human-as-a-Service (HuaaS):** Some principles look beyond the classical characterizations and introduce a new service level, the HuaaS. HuaaS is placed over the preceding three service levels and it means that the cloud paradigm is not just used for information systems, but also for humans [3]. Humans may be viewed as a resource because they have some capabilities which are more desireable than those of modeled information systems, for example creativity or translation tasks [3]. In the HuaaS, the dominating phenomenon is the so called ´´Crowdsourcing". It means, that quite complex tasks are split into granular activities, and those activities are distributed over the Internet to many unknown people [3], [12]. These very granular activities are paid only very little because they do not require a lot of experience or education. As we take a look at YouTube, which is typical for this genre, we see that every user of it can profit from the artifacts created by the whole community [3].

7.4 New Opportunities of Cloud Computing – What is it good for?

In the following section, several use cases are identified, where some kind of Cloud Computing Services are deployed. It will also be considered whether it is a pure Cloud Computing based approach or some kind of hybrid approach. Every use case shows the economic benefits for the company by using a specific Cloud Computing Services rather than a traditional IT infrastructure. Hereby traditional IT infrastructure means that all the IT requirements are managed by the company itself. In order to have a more accurate opinion, also some negative aspects will be considered. Some popular services which are offered today as Public Cloud Services will be discussed first.

7.4.1 Popular Cloud Computing Services

Due to its emerging state, the availability of Cloud Computing services is already quite numerous, as can be seen in the following graphic. The most popular services of every section will be briefly discussed.

Figure 7.4: Highlights of the Cloud Computing landscape [15]

This picture classifies these Cloud Computing Services in Enterprise-centric, Platform-as-a-Service, Web-centric and utility hosting. After a more detailed look at the description of these services, it turns out that these four terms classifying the Cloud Computing Services correspond to the basic definitions from the proceeding chapter. From every part of the initial definition, a typical one will be described more precisely. There will also be described why the naming of the picture fits into the initial definition.

- **Google App Engine:** With Google App Engine a user can run his self-developed web-application on the infrastructure of Google [13]. This means that Google offers webspace as well as the necessary development tools. Looking at the proceeding definitions, it becomes clear that Google App Engine is basically a PaaS (offering all the necessary tools for development), and this offering comes together via a Web-Service (also some kind of SaaS). It can also be used by a private user, as they offer free accounts [13]. For an SaaS example you may consider Google Docs, just to name one [3]. The access to this applications can be limited (for example for organizational members) or offered to the whole world. Google also offers the possibility to access ones applications transparently through ones web servers[13]. The billing is also shaped towards a full and flexible demand-oriented price plan, which means that one only has only to pay for the traffic which is in fact generated [13].

- **force.com:** Force.com is also a PaaS, but the service they offer is for more sophisticated software, which only enterprises would use. Also the various access methods may only have importance for large companies [21]. The aim is that a company must not bother with technology, so they can focus on their core business. The services offered are quite similar to Google App Engine, but the targeted groups are (large) companies [21].

- **terremark:** After a look at terremark.com it turns out that they offer IT infrastructure for companies, so they provide a kind of IaaS [29]. In contrast to the proceeding two service providers, terremark offers just the "raw" infrastructure without a clear software development procedure. The infrastructure from terremark covers computing, storage and network which can be down- or up scaled as needed. It is also possible to set up virtual servers [29]. They also ensure the highest standards in security are met, as it would be interesting for governmental organizations [10].

- **MOSSO:** MOSSO offers similar services as terremark, but not especially for companies. Private users or small businesses would also be in the right place if they used it [19]. MOSSO is an IaaS like terremark but the focus lies on target groups with smaller needs [19]. The billing is more fixed costs based, which means that when a server is never used, the service using entity will still have to pay for it [19].

- **Windows Azure:** With Windows Azure, Microsoft offers an operating system based on Cloud Services [18]. Windows Azure is situated in the center of the scheme, because it offers a little bit out of everything according to the Cloud Computing definition. With Windows Azure Microsoft is proposing some kind of a all-in-one solution. The service offers access to Microsoft's datacenter [6]. The whole system is running on the Windows operating system and Microsoft's software development tools like Visual Studio must be used to develop software in the .NET Environment [6]. Microsoft aims towards the goal that the software development process for Cloud Applications is similar to the traditional one [6]. This feature is also a bonus in usability.

After this brief presentation of some characteristic Cloud Computing services the great overlapping to the literature becomes clear. A first possible concern of Cloud Computing was that the offers would not be adequate for large companies. After this overview of the state-of-the-art in Cloud Computing services, this prohibiting anxiety can definitively be eliminated.

7.4.2 Cloud Computing – Use Cases

This subsection deals with three given scenarios in which some use cases of Cloud Computing are discussed. Foremost those scenarios are described in detail. After the background of each use case is established, it will be switched logically to the next section. The first use case describes a large company who owns and maintains its own e-Mail system. The second one is a company who runs its own bookshop and the third covers scientific computation. After the outline of each use case becomes clear, the following section contains a discussion about the economic challenges which continues the specific example given.

7.4.2.1 A company and its Mail Server

Let us assume a fictitious enterprise as described by the following: The enterprise "C.Laud-Serv. Inc." is a large business with about 10'000 employees. They are a leading company in the consulting sector and they serve clients from all over the world. Therefore, they need to run many branches in many different countries. The headquarter is situated in New York, USA. Until now, they run an E-Mail System by their own. That means, that every branch in every country has its own e-Mail server situated in the vicinity of the office. To achieve basic standardization in every branch they use the same settings and software, so that these can be considered a enterprise-wide de-facto standard. On each server, there are about 500 to 999 Mailboxes, which has been identified as average in a survey [23]. They are using on-premise e-Mail. This means, that all E-Mail Services are in the corporate data-center [23]. This is quite common for companies of large or middle size [23]. Running an on-premise e-Mail normally include the administration of the following [23]:

- Mail Servers
- Internal routing servers
- Mail Storage
- Storage archive, if necessary
- Client access servers
- Gateways
- Public Folders
- E-Mail Filtering

Even if it is not directly obvious, but a reliable and fast e-Mail System is vital for C.laud-Serv. Inc. Due to the large amount of customers they communicate with they create an enormous volume of electronic letters every day (just consider that e-Mail is becoming the most used communication tool in the business context). Since the personal fluctuation is quite high (let us assume that the average employee stays with C.laud-Serv. Inc. for a maximum of about two years) they put a lot of workhours into creating and deleting e-Mail accounts on the respective servers. In order to immediately recognize the origin of an enterprise intern e-Mail they used the following structure to create new e-Mail accounts: "name.surname@country.cloudServInc.com". Hereby, the country represents the origin country where an employee is hired. In the last two years, C.laud-Serv. Inc. acquired many new customers all over the world. At the moment they are thinking of raising the number of employees in the next few years to about 13'000 which would be an increasing by about 3'000 employees worldwide. But luckily a very vigilant employee from the IT-Infrastructure team mentioned that it may not be possible to follow through with these plans because most of the e-Mail servers are quite old (about five year in the average). But C.laud.Serv.Inc. is not the only company confronted with this problem as it is quite common that many firms run older e-Mail Systems [23]. Especially in the USA and in East Asia, where the biggest increase in customers is expected, they would need many brand new and powerful e-Mail servers in order to handle the new traffic which will be generated additionally every day by the large amount of e-Mail. The responsible manager from the division East Asia commissioned a little study in order to estimate the costs of such an upgrade of the e-Mail system. Hereby, there is a simple three step model of Forrester Research [23] available, which aids in the calculation of the costs which e-Mail are producing. The short description of this model is as following [23]:

1. **Estimating the e-Mail needs for each group of employees:** In general, there are three types of employees which have different needs in e-Mail. One segment is called **Mobile Executives**. Most of the time they are not in front of the desk (*e.g.* client meetings) and they need large mailboxes for the storage of big documents. An appropriate estimate is about two gigabyte per mailbox. The **Information workers** spend about one hour a day checking their e-Mails [23]. Therefore they need a mail client but smaller mailboxes of about one gigabyte [23]. Last but not least there are the **occasional users**. They don't use e-Mail so much, *e.g.* part-time workers or interns, so they need approximately around 250 Megabytes [23].

2. **Calculate the On-premise e-Mail cost:** The relevant cost calculation includes costs for maintainance and support for the e-Mail software, the costs for archiving and storing, staffing costs, hardware costs including power supply and financing costs for the e-Mail infrastructure [23].

   This cost estimating varies by the individual settings and is vaild for about 15'000 e-Mail using employees [23].

3. **Comparisation of the on-premise e-Mail against Cloud-Based alternatives:** The final step includes the valuation of other alternatives where the same cost position as in the on-premise e-Mail solution should be considered.

After the estimating process was completed the division manager was quite startled as he examined the amount this upgrade would cost. Many consultants also need large mail-

Table 7.1: Monthly cost per user using an on-premise e-Mail solution [23] (in US$)

Cost position:	**Ammount:**
Subscription	0.00
Server Hardware and OS	0.56
Server Software	3.61
Client Software	3.49
Storage	1.23
Message filtering	2.99
Message archiving	8.89
Staffing	4.41
Total Ammount:	**25.18**

boxes because they are dealing with many different customers at the same time. But recently, he had heard about Cloud Computing Services, which could also be used for the handling of e-Mails. The main triggers for migration to a Cloud Service solution are e-Mail getting too expensive, consuming IT Staff time and resources for the maintenance process, a recent survey of about 53 firms shows [23]. As Cloud Computing as an emerging technology could reduce the cost, he started to inform himself about the available possibilities.

7.4.2.2 A company and its Bookshop

The following use case is considering a company which plans to set up its own book shop. As all the necessary precautions and preparations on the logistic part (i.e., what books and where they can be bought) are done, there will be the question how to concretely implement this book shop. In order to deliver the books, they have already established some kind of just-in-time delivery, which allows the incoming books to rapidly be shipped to the customer. Firstly, the bookshop "Book-C" was founded as a little startup by two people, Mark and Jack, which at the beginning offered books only to the people which were personally related to the founders of this particular bookshop. This means, that at the beginning the access rate to this bookshop was quite low. After the first year, they could acquire quite a number of clients who access this bookshop more or less regularly, so that this online shop can not be operated any longer on the owners' personal computers. According to their predictions, the numbers of customers who access the shop more or less regularly may raise from 100 accesses a day today to more than 1000 accesses a day in a one year period. After that it would slowly raise to more than 10000 accesses a day, with the access peak in the evening. Now, in order to guarantee the correct operation and accessibility of this book shop in future, they consider several solutions for their future investments. On one side, there is the traditional possibility to buy new servers and operate the whole IT Infrastructure on their own. For many years, Mark worked before in the support department of a large IT company and he is quite familiar with IT infrastructure and service. Therefore he would prefer the traditional solution which means the operation of own servers. But Jack, who has just finished his studies, does not have any experience in the area of IT infrastructure, therefore he would prefer the choice of a cloud computing solution.

In order to keep this use case from growing too complex, let us consider that they use a simple database in which all the necessary shop information are stored. Further, there will be a web interface on which the customers can interact with the bookshop's database. Therefore, viewed from a traditional point it would be necessary to set up servers for the web interface and for the data base. As the access rate is expected to grow constantly, there will be needed approximately ten servers in order to guarantee a fast and reliable access even at peak time. They have to be fast, as well as equipped with high storage capacity, because the target of sellable items of this bookshop will be around 1 million in about five years. As this large amount of items will require much storage for the database, the question of storage will be quite important in the future. Furthermore they intend to host the online shop for books, which means they not just ship the books to the customer, but also offer books in an electronic version, which can be downloaded to the customer's personal computer. This requires the implementation of a billing system which can be upscaled for the future.

7.4.2.3 Scientific computation in the cloud

Research groups in the area of applied sciences often need raw CPU power to conduct specific calculations on scientific models. The models are often based on complex formulae so that only one workstation or server is not sufficient to do such calculations in a reasonable time frame. This matter of fact lead to different approaches, from building time-shared local cluster systems as far as inter-university based grid systems. Depending on the monetary involvement, those resources are in an available in a limited way. Even if there are urgent time-slots on clusters/grids, it cannot be guaranteed that a researcher will have the calculations within his required time frame. Those facts could lead to the conclusion that a research group or an institute of a university would need its own computational resources. This example is based on the following hardware requirements. The computational "cluster" consists of four nodes where each node has eight (virtual) CPU cores at 2.5Ghz, 6GB Ram and about 1.5TB hard disk space. The executed calculations will result in a relatively high resource usage of 40% overall. One way to get the needed computational power would be to build system and just use it like it has been done in the past. Since the emerging area of Cloud Computing allows to use and share processing power in a smarter way, it is more likely to run on the cluster an implementation software for Cloud Computing. Such an example would be Eucalyptus [30], which is the abbreviation for "Elastic Utility Computing Architecture for Linking Your Programs to Useful Systems". Since one of the main features of Eucalyptus is the compatibility with the Amazon Web Services API, it is obvious to compare the prices in this use case for a self-managed cloud based on the cluster system and just using the Amazon EC2 services.

There is still one very important aspect of using an external cloud, even if all the security mechanisms are well known and established. The most valuable assets of applied science research groups lie in the research itself and the gained insights. Sometimes there are also collaboration projects with private enterprises. This fact makes the area of secure data access and exchange especially important. Besides the well known approaches in the area of identity management, secure cryptographic key management and access control policy enforcement, the question of outsourcing research critical data to a commercial cloud service provider is most essential. Since this discussion would blast this use case, the basic assumption is to move only the raw computation to the cloud. This would also allow using the commercial providers, since the delivered security mechanisms are adequate and the main data mining and conclusion composition will take place at the university networks.

This finally leads to the scenario where the institute needs to maintain a specific amount of CPU power. For the sake of example to compare a self-maintained cloud solution and Amazon EC2 reserved instances, the time horizon of usage will be three years.

7.5 Economic Aspects and Challenges of Cloud Computing

The following section has the aim to identify economic challenges and the underlying economic principles which prevent large enterprises to employ Cloud Computing. In this section the previous examples are continued and discussed in more detail, especially concerning what the possible advantages and disadvantages for a solution would be. Now, as a general understanding of Cloud Computing is being set up, a further look at the economic challenges can be done. Hereby the proceeding use cases are followed and expanded with further details.

One side are the challenges derived directly from the use cases. On the other side they are more general issues which are obstacles in implementing Cloud Computing Services.

7.5.1 Challenges in the first use case – The e-Mail system

As we see in the table below, at a level of about 15'000 employees the cost of an on-premise solution is as high as a cloud-based one [23]. Therefore, every company smaller than 15'000 employees, that use e-Mail according to our assumptions should consider moving to a cloud-based solution. At the beak-even point of about 15'000 employees it turns out why small companies often easily switch to a cloud based solution as one of the assumption of this seminar proposes. This can be explained by examining the cost structure of an IT-Infrastructure: Generally in economy, costs can be separated in fix-costs and variable-costs and it can be derived that setting up and maintaining an e-Mail server has quite a large fix-cost component in it. This fix-cost remains the same no matter if only 1 employee uses it or 500. When an additional server has to be set up, than the fix-cost will raise again. For the whole calculation this fix-cost can be divided by the number of employees using e-Mails. Generally, when the number of employees is higher this component is decreasing. As we have shown in this paper, using Cloud-Computing services almost always results in a very small fix-cost component (*e.g.* a subscription fee or something similar). In every case this is much smaller than the fix-cost for an own e-Mail server farm. And at all time only the actually used resources have to be paid, these are the variable costs. Thus this model is especially attractive to small companies which do not have many employees. They can profit from paying only the variable costs, which are much lower. As we seen in the following table, with about 15'000 employees paying 15'000 times a variable cost component, it becomes equal to paying a fix-cost component for 15'000 employees. With a number of more than 15'000 employees the variable costs are rising faster than the own fix costs for on-premise e-Mail server.

Table 7.2: Monthly cost per user comparating On-premise, Cloud-based, Microsoft Exchange and Google Apps [23] all the prices are held in US$

Cost position:	**On-premise:**	**Cloud-based:**	**Exchange:**	**Google Apps:**
Subscription	0.00	9.78	8.66	4.17
Server Hardware and OS	0.56	0.00	0.00	0.00
Server Software	3.61	0.00	0.00	0.00
Client Software	3.49	3.49	3.49	0.00
Storage	1.23	0.00	0.00	0.00
Message filtering	2.99	1.86	0.00	0.00
Message archiving	8.89	8.11	6.33	3.75
Staffing	4.41	1.85	1.85	0.55
Total Ammount:	**25.18**	**25.08**	**20.32**	**8.47**

In the proceeding table the low cost of Google Mail [14] is standing out. It is reasonable to assume that Google can offer this service because it is intending to offer its service to many many users and therefore the fix-cost for this e-Mail farm can be divided by a large number of customers. Also in case of a large percentage of mobile workers the opportunity to access the e-Mails in the cloud via mobile devices would be a great advantage. Because of these points, it would be reasonable for the C.Laud-Serv. Inc. to move to a cloud-based solution.

7.5.2 Challenges in the second use case – The book shop

Choosing the traditional solution, which means operating and maintaining the servers at their own will be a quite high cost factor. For example let us assume that a good server costs about $ 3,500 [9]. As the company "Book-C" will buy ten of these servers, they already would have to pay $ 35,000. As both Mark and Jack live in small flats, they don't have the opportunity to operate the servers in their rooms, therefore they have to rent another location for the server wich will cost about $ 500 a month (assuming this amount to rent a flat for an office in quite unattractive location). The cost for electricity in order to operate the servers will be about $ 100 a month and maintenance is estimated at $ 500 a month (as the maintainings can be done by Mark). Also, they optimistically think that their servers will run for about 7 years, which means that every year their little server farm will devalue about $ 5000 every year (35000 / 7 = 5000). So, they have to compensate this amount of $ 5000 every year in order to be able to buy new servers after the period of 7 years (they assume, that with a total amount of $ 35000 there could be satisfied the higher access rate, because it is likely to expect that the hardware cost will be lower). As we have seen so far, the total cost will be about $ 35000 which will be paid right now and another of $ 1100 for the monthly mainting and operating costs and a rounded up amount of $ 420 for the depreciation. For a fast internet connection there will be about $ 80 in charge for one month. This will lead to a total ammount of $ 1600 per month. Comparing this with the cloud based solution, there will not be any charge for the individual traffic, which means that the cost for the internet connection will always be $ 80 per month, no matter if 1 person or 1000 people access the shop.

As Jack was quite familiar with the Amazon Web Services [1] they also did a cost calculation for the cloud computing solution which will be operated with Amazon Web Services. As it turns out clearly, there will be no fix cost, this means, they don't have to pay the $ 35000 in order to buy the servers. In addition they do not need an extra office location and they will not have to pay for the electricity. Also, the servers do not need to be maintained by Mark, which saves another $ 500 per month. They also don't have to deal with the depreciation of the servers because they don't bought any. But what kind of cost component will be here? The answer is simple, there will only be variable costs. In this example there will be costs for the storage of the files and for the traffic from and to the servers. Also a big advantage of using cloud services will lie in the flexibility of upscaling the whole service [1] to higher access rates. As this is expected to become necessary in the future, the possibility of easy upscaling will count as a great advantage. As we consider the cost for storage in the Amazon Web Services, the cost of storage for up to 50 terabytes will be 0,150 USD / GB [1]. For example let us assume that they will need around 100 GB every month, this will cost about $ 15. As we see, the cost is much lower in this case. Therefore it should be chosen a cloud based solution in order to establish a bookshop.

7.5.3 Challenges in the third use case – Scientific computation

As mentioned before, this use case mainly covers the economic aspects between using Amazon EC2 as IaaS provider against deploying an own cloud environment. The calculations are done for four nodes during a usage time window of three years. The average resource usage will be at 40% during this time. Further storage possibilities on Amazon Web Services like using Amazon S3 storage will not be investigated, since the focus lies on raw computational power.

The following table shows the calculated costs for both options. Since all nodes are running on open source software, no further license costs are accumulated. To have test best comparison, EC2 Reserved Instances will be used rather than on-demand instances. The reserved instances are more suitable for the scenario and based on the expected usage, they will result in lower total costs even if there are initial fix costs involved. The used instance type for EC2 are four "High-CPU Extra Large (XL) Instance" packages [31] located at the Ireland data center. The example hardware for self deployment is calculated by the costs of hardware similar to the Amazon solution and the electricity costs. The electricity costs per node are calculated with the current price per kWh in Zurich, averaged to a 24/7 tariff of 0.1415 US$. One node consumes approximately 0.5kWh. Since most universities usually have their own data centers and IT services, those costs are ignored in the self deployment scenario. One aspect to mention is that the self deployment scenario hardware specifications are slightly superior to the EC2 configurations.

Table 7.3: Amazon EC2 versus own infrastructure costs for three years of usage

Cost position:	**Own infrastructure:**	**Amazon EC2:**
Initial fixed costs	12'500	11'200
Monthly variable costs	206	420
Lifetime variable costs	7'416	15'120
Lifetime total costs:	**19'916**	**26'320**

The table shows that for the example use case the deployment of an own cluster based cloud will save 6404 US$ over the whole life time period of three years. This leads to the conclusion that in an environment where the resource usage is pretty high, the self deployment solution is mostly superior. If there is less resource usage, the scenario has to be reevaluated. If a further rise of resource usage is expected, the self deployment scenario will result in lower costs. Another advantage is, that there are nearly no costs per unit for an additional virtual machine instance in the self deployment scenario. This would allow migrating further services from physical hosts to the cloud. Surely the self deployment scenario requires more knowledge in IT infrastructure but besides the financial benefits, it allows to handle research critical data in-house to ensure confidentiality.

This use case has shown that there is not a simple question of using Cloud Services or not. Each time the required resources have to be determined before any decision can be made. Having the expected resource usage and plans for further expansion, the suitable Cloud Computing strategy can be chosen, for example IaaS. Sometimes, especially in the area of handling critical and/or confidential data, outsourcing to a Cloud Service provider is no option. This does not mean that such an area is prohibitive to Cloud Computing, even more the idea of creating an own Cloud can provide the desired cost and operational efficiency. This allows to manage access control and identity management in-house without having critical data leaving the own network. An additional benefit of a self deployed Cloud environment based on open source software is the fact, that there is no vendor lock-in. Such a lock-in would result in a dependency to a specific service provider, which is neither in academics nor in business a wished effect on the customer side.

7.5.4 General Identified Issues

The general identified issues are explained with the principles of the New Institutional Economics (NIE) as well as with general security issues.

7.5.4.1 Transaction cost theory and opportunistic behavior

One fundamental economic aspect which prevent Cloud Computing for being implemented in especially large companies could be explained with the transaction cost theory and in particular with a opportunistic behavior [4]. Because of the partly or complete outsourcing of IT-Infrastructure, services or the platforms a company loses the control over these parts. As we have seen before, using services of Cloud Computing means always that these services are not any longer in the proper control, then, they are under complete control of the operator. In other words, a company who decides to use cloud computing services does not have to deal any longer with this part of IT which has being outsourced to the Cloud Company provider. But form this point on, instead of dealing with these IT matters, the company has to deal with the provider, also, it has to handle these transactions. Therefore, the transaction cost theory seems to be a good model to explain this fact. Nowadays, in numerous companies the IT has a fundamental value. Banks without IT would not be possible to work any more [24]. Therefore, when a company uses some kind of Cloud Services, it could being blackmailed by the provider due to this loss of control over its own IT [4]. This means, that a provider could willingly raise his prices for several services and due to the well-known lock-in effect the company can not change. A way to reduce this could be a high standardization of the underlying components [11].

7.5.4.2 Security – A cloudy outlook?

- **Data Theft** Nowadays several case have happened where critical data was stolen and sold afterward to governmental institutions [28]. As we have seen before, a typical Cloud service (the IaaS one) offers the whole infrastructure to store foreign and client data. Depending on the data a company possesses, the risk of giving the data out of its hand can be very high. A consulting company which has only immaterial products (which are stored and managed on IT systems) would surly have to declare bankruptcy if their data had been stolen or lost. Hereby the risk of a data loss should be very low, because every of the introduced Cloud Service providers ensure that the data are very well back-upped and safe. But it is always difficult to judge whether the data is really stored as safe as it was announced. In this case, maybe choosing just an SaaS would be appropriate? Furthermore a consulting company may not want that their own software is given out their hands.

  Even if the provider does not follow opportunistic behavior, it could be possible that he is forced to raise the prices, because of other circumstances. In such a case, a company would have to accept it, because Cloud providers often use proprietary environments [20] (as it was mentioned with the PaaS).

- **Attacks over the internet** Another possible scenario proposes that the datacenter itself is attacked by hackers. Especially if it becomes public knowledge that several large and important companies are using the same Cloud Service provider to store their data, this Service provider may become a very very attractive aim for crackers. The global availability of such a Cloud Service provider even increases the attractivity for intruders. Even if almost every commercial Cloud provider claims to have implemented very high security standards [13], [18], [29], [19], [21], it is still possible (but with a very low probability) that data can be stolen by attackers. In any case, the risk has to be considered high, if we calculate it as the product of the incident rate and the size of the loss. The possible damage would be very high for a specific company, because all of its data would be stolen and clients compromised.

7.5.5 Preliminary solutions to the identified challenges

As Cloud Computing is still lacking proper standardization lock-in and opportunistic behavior is a big strategic obstacle. Unfortunately the solutions that can be advised today all go together with higher expenses. One the one hand it is advisable to store data at different locations in order to limit the extend of opportunistic behavior. This is easily achieved by keeping backups locally or at a different provider, that is not connected to the Cloud Service Provider.

If one really needs to store critical data off-premises or offshore make sure to anonymize and split it as much as possible to reduce the value of the data for any intruder. For example book orders could be identified by a unique transaction number in the cloud, but the mapping which transaction number belongs to which client should be stored in ones sphere of influence. Additionally when dealing with sensible client data it is worth researching where the data is actually going to be stored geographically. This directly influences the risk of data being accessed through legal means, that would have been unavailable in other parts of the world.

Another problem is the physical dependency from the Cloud Provider. If a Cloud Provider is hit by a outage or disaster your own business might be in risk too. This problem does not really have to do with the technology the provider uses, but has to be seen as a general problem of specialization. Due to the very generic nature of this thread there is no popper solution, just a basic guideline. Always try to pick established and solid partners if your business depends on the availability of the service provider. This is most important for big companies and can be a prohibitive obstacle in implementing Cloud Computing. Smaller businesses do not face this thread as severely as they are more agile and could switch more rapid to another provider if necessary.

7.6 Conclusion

Cloud Computing is a powerful tool that can enable small companies to serve large user groups on a flexible basis. But as we have illustrated Cloud Computing does not only offer benefits to small enterprises, but also offers many possibilities for bigger enterprises that want to focus on their core capabilities.

Cloud architectures can be categorized as being public – when the service is offered to virtually anyone –, as private – where the provider and user reside in the same organzation – or as hybrid – where some data is kept local and some is transfered to public clouds. The service types that build on top are SaaS, PaaS and IaaS – each increasing the control and workload for the client gradually. Where SaaS is considered a solution that just works for a specific task, PaaS just offers the tools to create such a solution and IaaS just provides the hardware.

It has to be decided on a "‘case per case"’ base whether it is beneficial to outsource certain services to the Cloud. One of the main advantage is that support and maintenance of the hardware is given away, which enables scientists and non IT companies to receive a higher level of reliability compared to self-hosting.

Problems arise when companies only see the good and miss the predicaments of Cloud Computing. Security often comes to mind first, and companies will have to realize which data can leave their servers and which is too critical to trust anyone else with it. One not only needs to aim to find a fair balance between kinds of data, but also for security itself. A professional Cloud Provider might offer better protection than your own network, *e.g.* due to a dedicated security team, but on the other hand a Cloud that stores important information of different companies becomes a valuable target for attackers.

As for now Cloud Computing does lack basic standards, which makes it hard to switch a medium or big company from one provider to another. The lock-in this provides is pretty strong and it might hinder many companies to transfer their data to the cloud. Because of this Cloud Computing should only used for basic or auxiliary services at the time being.

Bibliography

[1] Amazon Web Services: `http://aws.amazon.com/de/`

[2] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. H. Katz, A. Konwinski, G. Lee, D. A. Patterson, A. Rabkin, I. Stoica, M. Zaharia: *Above the Clouds: A Berkeley View of Cloud Computing*, Electrical Engineering and Computer Sciences University of California at Berkeley, Technical Report No. UCB/EECS-2009-28, 10 February 2009.

[3] C. Baun, M. Kunze, J. Nimis, S. Tai: *Cloud Computing*, Springer Berlin / Heidelberg, 2010.

[4] C. Breuer, W. Breuer: *Shared-Services in Unternehmensverbuenden und Konzernen: Eine Analyse auf der Grundlage der Transaktionskostentheorie*, Gabler, 2006.

[5] K. Bennett, P. Layzell, D. Budgen, P. Brereton, L. Macaulay, M. Munro: *Service-Based Software: The Future for Flexible Software*, Seventh Asia-Pacific Software Engineering Conference (APSEC'00), 2008.

[6] D. Chappell: *Introducing the Windows Azure Platform*, David Chappell and Associates, December 2009, `http://www.microsoft.com/windowsazure/whitepapers/`

[7] Cloudhostingprovider.co.uk: *Basic Knowledge about Cloud Computing*, 26 December 2009, `http://cloudhostingprovider.co.uk/`

[8] E. Deelman, G. Singh, M. Livny, B. Berriman, J. Good: *The cost of doing science on the cloud: the Montage example*, ACM/IEEE Conference on High Performance Networking and Computing, IEEE Press, Austin, Texas, USA, November 2008.

[9] digitec: *Marken-Server* 10 April 2010, `www.digitec.ch/`

[10] J. Staten, L. E. Nelson, A. Herald, S. Yates *Case Study: USA.gov Achieves Cloud Bursting Efficiency Using Terremark's Enterprise Cloud*, Forrester, 25 September 2009.

[11] Forrester Consulting: *Supporting Sustainable Cloud Services: Investing In The Network To Deliver Scalable, Reliable, And Secure Cloud Computing*, A commissioned study conducted by Forrester Consulting on behalf of Juniper Networks, October 2009.

[12] O. Gassmann, E. Enkel: *Towards a Theory of Open Innovation: Three Core Process Archetypes*, Institute of Technology Management, University of St. Gallen, Switzerland, 2004, http://www.alexandria.unisg.ch/Publikationen/274

[13] Google: *What is Google App Engine?*, Google Code, http://code.google.com/intl/de-DE/appengine/docs/whatisgoogleappengine.html

[14] Google: *Google Mail für Unternehmen* http://www.google.com/apps/intl/de/business/googlemail.html

[15] D. Hinchcliffe: *Cloud computing: A new era of IT opportunity and challenges*, ZDnet, Enterprise Web 2.0, 3 March 2009, http://blogs.zdnet.com/Hinchcliffe/?p=261

[16] IBM: *Cloud computing: Paradigm shift or just hype?*, Express Advantage, https://www-304.ibm.com/businesscenter/cpe/html0/158782.html

[17] C.-P. Tiemann: *Messe in der Krise, die IT boomt*, Limmattaler Zeitung, LIZ, p.10, 02 March 2010.

[18] Microsoft Corp: *Windows Azure*, Windows Azure Platform, http://www.microsoft.com/windowsazure/windowsazure/

[19] MOSSO: *Welcome to the Rackspace Cloud*, http://www.rackspacecloud.com/

[20] Oracle *Architectural Strategies for Cloud Computing*, An Oracle White Paper in Enterprise Architecture, August 2009.

[21] Force.com: *The leading cloud platform for business apps*, salesforce.com, http://www.salesforce.com/platform/

[22] L. Qian, Z. Luo, Y. Du, L. Guo *Cloud Computing: An Overview*, Springer Berlin / Heidelberg, 2009.

[23] T. Schadler: *Should Your E-Mail Live In The Clouds? A Comparative Cost Analysis*, Forrester, 05 January 2009.

[24] G. Schwabe, P. Nussbaumer *Why IT is not used in the Advisory*, The 17th European Conference on Information Systems (ECIS 2009), 2009.

[25] J. H. Scanlon, B. Wieners: *The Internet Cloud*, The Industry Standard, 05 March 2010, http://www.thestandard.com/article/0,1902,5466,00.html

[26] D. Smith: *Google plans to make PCs history* The observer, 25 January 2009, http://www.guardian.co.uk/technology/2009/jan/25/google-drive-gdrive-internet

[27] Stadt Zürich, Finanzdepartement: *IT Strategie: Umsetzung der neuen IT-Strategie der Stadt Zürich zur Eindämmung der Informatikkosten*, http://www.stadt-zuerich.ch/fd/de/index/informatik_oiz/IT_Strategie.html#ziel_6_reduktionderbetriebszentrenaufzweigrosserechenzentren

[28] Tagblatt: *Datenklau: Welche Banken sind betroffen?*, Tagblatt online, 03 February 2010, http://www.tagblatt.ch/aktuell/wirtschaft/wirtschaft/Datenklau-Welche-Banken-sind-betroffen-;art623,1486967

[29] Terremark Worldwide Inc.: *Leading IT Infrastructure Service*, The Enterprise Cloud, http://www.terremark.com/services/cloudcomputing/theenterprisecloud.aspx

[30] Eucalyptus Systems: *Eucalyptus | Your environment. Our industry leading cloud computing software.*, http://www.eucalyptus.com/

[31] Amazon Web Services: *Amazon EC2-Instanztypen*, http://aws.amazon.com/de/ec2/instance-types/

Chapter 8

Smart Meter

Reto Seeholzer, Maurice Göldi, Igor Bozic

According to the ordinance of the EU, every household should be vested with a smart meter. Smart meters record exactly, nearly every second, the power consumption of a household. To determine the device that uses most energy in a household and being able to adapt the behavior, an exact feedback of the power consumption is necessary. New devices like smart meters allow a detailed note of the complete power consumption of a household. A smart meter measures the quantity of the energy and the time the energy was consumed. It stores the data electronically and transfers it to a data collector or utility. Smart meters alone don't save energy, it is just a device that shows the actual consumption and out of that the household can reduce the energy consumption and save energy. The energy usage should be put at the time when the offer is greater than the demand. The goal is to reduce the power consumption.

Contents

8.1 Introduction

A smart meter is a meter that determines and stores in real-time or near real-time energy consumption. Smart meter often refers to an electrical meter, but it can increasingly also mean a device, measuring water or gas consumption. They have a great potential to save energy. The data, of the actual energy usage, will be delivered in short intervals through the internet to the particular provider. This is why smart meter allow a monthly bill of cost, because of the actual consumption. Smart meter allow dynamic or new tariffs, which gives new incentive to save the energy consumption [1].

Smart meter allow a detailed measurement of a usage like electricity. With this kind of information's, consumers can control and maybe they can save energy. Providers on the other side can make variable tariffs, at different times they can offer particularly favorable prices. Usually meter readers are installed in the cellar, and people don't pay attention to them. They show only the total consumption but they don't give any information about the current usage. But smart meters have helpful advantages, they inform the user about the current energy that is used by a household and they can help to be economical. A smart meter is a new type of tool to measure and record how much electricity you use at different times of the day and week. A smart meter measures the quantity of the energy and when was it consumed. It stores the data electronically and transfers it to a data collector or utility. It can facilitate various tariff options e.g. time of use, time of day, rising block etc.

The goal of a smart meter is that the costumer should reduce his power consumption. The energy usage should be put to a time when the offer is greater than the demand [2].

8.2 Technology

8.2.1 Terminology

The technology for metering is undergoing a significant change at the moment. Due to new laws in most European countries the classical electromechanical meters with manual meter reading are being replaced by electronic smart meters with automated meter reading. The literature identifies four different types of systems [4] [5].

- Advanced Meter Reading (AMR)
- Advanced Meter Management (AMM)
- Advanced Meter Infrastructure (AMI)
- Smart Metering (SMET)

8.2.1.1 AMR – Advanced Meter Reading

Advanced Meter Reading describes systems that allow a remote meter reading. This technology is already a standard for industry consumers in Germany. But while here the data is read out by modem once a day, the implementation for private customers allows for different communication systems to read out the data:

- Meter reading directly at the meter through an interface with a hand held device.
- Meter reading through a local communication system (usually wireless transmission) so that the building doesn't have to be entered. Data can be collected in a walk-by or drive-by manner.
- Meter reading through a long distance communication system. This makes loca meter reading obsolete.

8.2.1.2 AMM – Advanced Metering Management

In contrast to the AMR systems, an AMM system allows for bi-directional communication either through a separate or integrated communication gateway. This on one hand allows for read out of meter data and on the other hand pricing information or a control signal can be transmitted to the meter. This allows for additional functions in the smart meter that can benefit the consumer.

8.2.1.3 AMI – Advanced Metering Infrastructure

AMI is an expansion of the AMM. It includes the extended functions needed for grid control (smart grid) and intelligent house controls (smart home) and meter-data-management systems.

8.2.1.4 Smart Metering

Smart metering is usually used as an umbrella term for the concepts mentioned above. A system composed of meters, a communication system and a meter-data-management system (MDM) is referred to as a smart metering system.

8.2.2 Communication Interfaces

A smart metering system in principle has the following functions: count/measure, data acquisition, save and control data, communicate and additional value functions. Communication is the essential function of the smart metering system. In general all meters have a local communication port to read out data. Additionally they have a bi- or unidirectional communication port for long range data transmission. Some systems have additional ports to integrate other meters (e.g. gas, water, thermal energy) and to interface with customer applications (e.g. displays). Some systems also allow data transfer from a central point to different service providers.

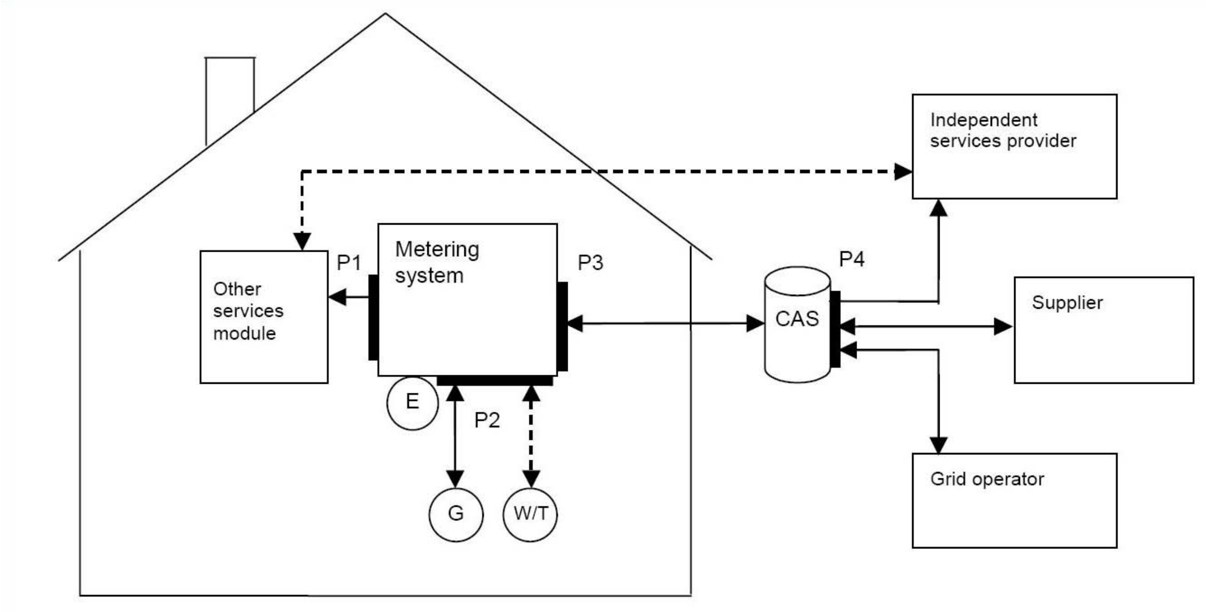

Figure 8.1: Communication ports belonging to the metering installation

As well as the various displays on which the actual meter readings can be taken, the metering installation has the following communication ports [6] [7]:

- port P1 for the communication between the metering installation and one or more other service modules, but it cannot be used for sending data to the metering system
- port P2 for the communication between the metering systems and one up to four metering instruments and/or grid operator equipments
- port P3 for the communication between the metering installation and the central access server (CAS)
- port P4 is the port on the CAS with which independent service providers, suppliers and grid operators gain access to the CAS

The communication via ports P1 to P4 are handled by wireless or wired communication technologies. For the long range communication of data via P3 there are two different concepts used.

- Direct data transmission (Point-to-Point, P2P): The meters are directly connected to a central server. Mostly wireless (GSM/GPRS) and wired technologies (DSL, Fiber optics or TV-Cable) are used!
- Indirect data transmission (Point-to-Multipoint, P2MP): The meters of multiple houses (customers) transmit their data to an intermediate station where the data is integrated and then transmitted to a central server. RF-transmissions and Powerline Communication (PLC) are the most frequently used technologies here.

8.2.3 Hardware Concepts

The requirements of metering, data acquisition, control, communication and additional value functions can be implemented with different hardware concepts. The simplest solution integrates all these functionalities in one piece of hardware. Other solutions allow the individual components providing different functionalities to be added via a gateway. The hardware implementation of the smart metering system defines what kind of services can be provided, and if these services can be offered by various service providers [4].

8.2.3.1 Integrated Meter

The integrated meter contains all hardware necessary for metering, data acquisition, control and communication in a single device. The individual components cannot be exchanged. This effectively means that, for example the mode of long range data transmission, is already set on installation and cannot simply be changed.

8.2.3.2 Partly Modular Meters

In the partly modular meters the component for the long range communication with a service provider is modular. This means the component can be upgraded at a later date, the mode of communication (PLC or GPRS) can even be changed. This gives the consumer and the service provider more flexibility in the technical systems he intends to deploy.

8.2.3.3 Modular Meters

The modular meter doesn't only offer modular connections for the long range components, but also allows the short range communication and added service value components to be exchanged. While this is the most complex setup it allows the user the greatest flexibility.

8.2.3.4 Electronic Meter

The simple electronic meter only has the ability to measure the consumption of a resource such as electricity, water, gas or thermal energy. It then immediately transfers its data, through the local communication port 1, to a gateway which then houses the hardware to further process the information. There are also slightly more sophisticated models that have a display that can show the consumer historic values like yearly, monthly, weekly and daily consumption.

8.2.3.5 Integrated Gateway

The integrated gateway separates the storing, communication and control of data from the actual metering on a hardware level. The metering hardware for various resources can be connected through the local communication port, and the acquired data is then processed in the gateway.

8.2.3.6 Modular Gateway

Just as with the full metering systems there is a modular option for the gateways. Again the short and long range communication and the added value service components can be exchanged in a modular fashion due to standardized sockets, offering the service provider and the customer a maximum of flexibility. This hardware system has been defined but there are no available products on the market as of yet.

8.2.4 System Concepts

Another way to differentiate between different smart metering systems is to look at the system-wide concept employed. In particular how the sub-components metering, long range communication and centralized metering-data-management system (MDM) interact with each other. Currently there are the following systems on the market.

8.2.4.1 Open Central

In this concept the MDM-system can integrate data from communication systems of various manufacturers. This requires that the manufacturers of metering systems make their protocols available.

8.2.4.2 Central + Communication System

This concept has a centralized MDM-system and data concentrators or gateways. Meters for various resources can then be connected to this system through local communication ports. The long range communication is sometimes proprietary which means other concentrators or gateways cannot be integrated.

8.2.4.3 Communication System + Meter

This concept bundles the other end of the smart metering system. Data concentrators or gateways that can integrate data from multiple customers and for various commodities from the meters on one side and the central on the other. Here the first part of the communication, between meter and gateway or integrator, is generally proprietary. The second part to the central is then open for system integration.

8.2.4.4 Fully Integrated Smart Metering System

In this system all components from the meter to the central are run in a closed proprietary system. Although some systems allow foreign meters to be integrated through special gateways, the usual case does not allow this. This is currently the most widely used system variant for smart metering concepts.

8.3 Market Forecasts

Today there exist two different situations, households with a smart meter and households with a normal standard meter. Figure 8.2 shows such a scenario.

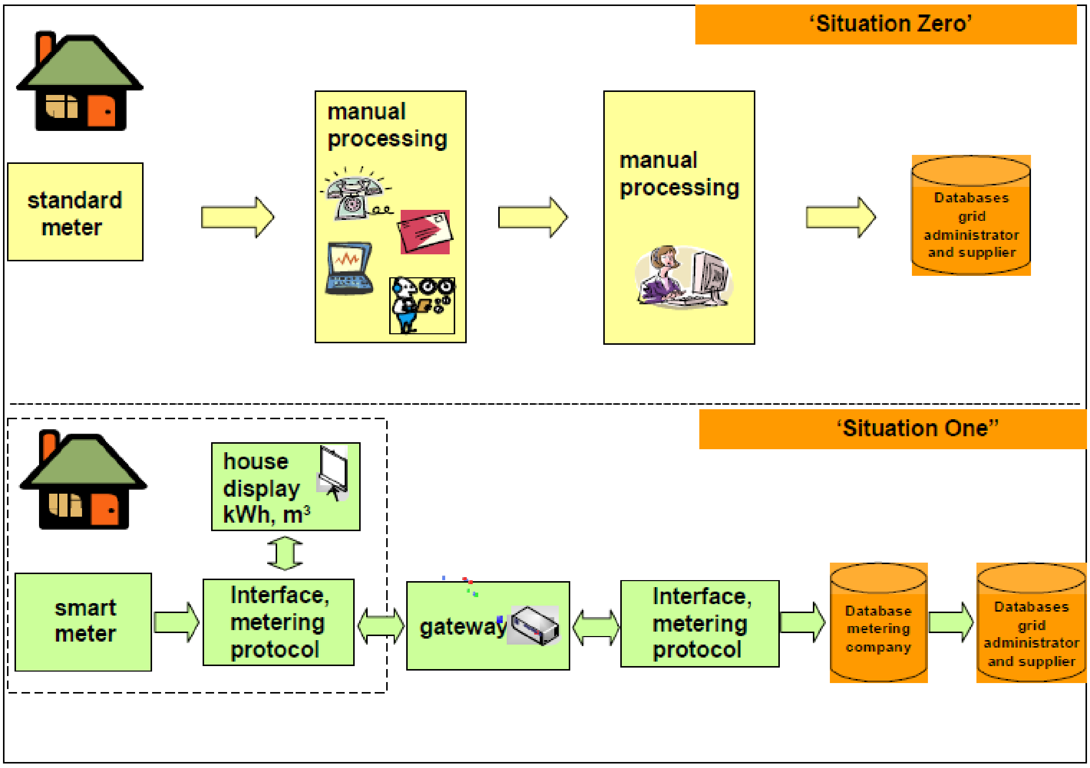

Figure 8.2: Situation Zero and Situation One [8]

- Situation zero is the current situation in which households do not have a smart meter and are not connected to a smart metering infrastructure
- Situation one is the situation in which a household is connected to a smart metering infrastructure and these households receive monthly feedback about their actual consumption

The transition phase from situation zero to situation one will be realized by the market players and according to the ordinance of the EU, every household should be vested whit a smart meter. They will ensure that meters and smart metering infrastructure will have an open structure so that consumers can change suppliers easily without any problems [8].

8.3.1 Variable Tariffs

Through the usage of a smart meter there will be new tariffs and new services. Time varying tariffs, e.g. peak related increasing block or seasonal. Through a smart meter it can be shown to the consumer, that he should use his washing machine at off-peak times. Household should use energy at times where the price is at lowest point and not at peak period. But new tariffs are not just only a benefit, they could create complexity and scope for confusion e.g. are the mobile phone tariffs, which are offered by the different mobile phone providers.

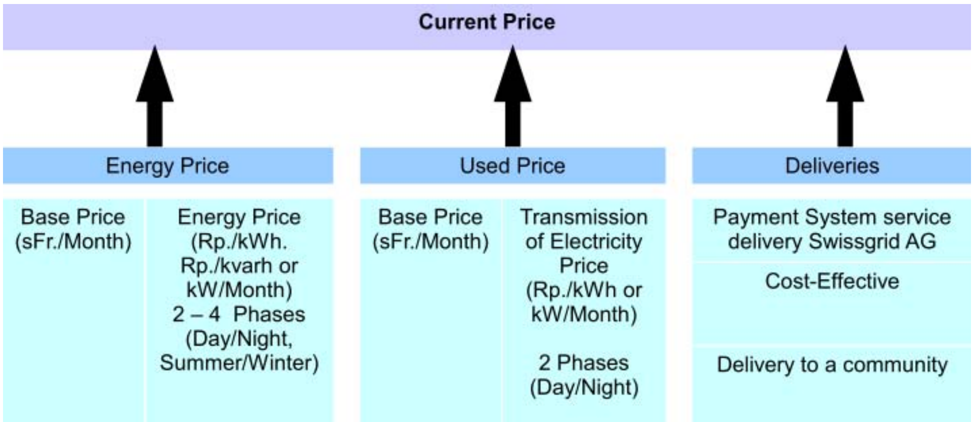

Figure 8.3: Structure of electricity tariff [4]

8.3.2 Time of Use Pricing

Time of use pricing, critical peak pricing and real-time pricing have been major concerns mostly in those parts of the world with summer and winter peaks in demand allied with supply constraints. There is evidence of reductions of up to 30% of peak demand through a variety of reduction programs involving some sort of time-sensitive pricing, with or without direct load control by the supplier. Automatic Meter Management allows for time-sensitive tariffs to be used and communicated remotely to the consumer. There is also a danger that it would penalize some low-income householders with lifestyles that make it very difficult to alter their consumption patterns. This picture could change as the proportion of demand that is met by building-integrated renewable and distributed generation increases, but that is a subject for further research. Electricity tariff structures do not have to be based on time of use in order to have an impact on load: for example, progressive block pricing offers an incentive to conserve and could be combined with informative billing and displays of how close the customer is to reaching a threshold above which the unit cost will be higher [9]. Today, under the current system, the prices for electricity are the same every day and every hour of the day. With the installation of a smart meter, the price of electricity will depend on when the energy is used.

- The usage of the dishwasher should start after 9:00 p.m. when off-peak prices begin, if the dishwasher has a timer, the timer should then be used
- Dryers consume a lot of energy. It should be used at the evening or the weekends

8.3.3 Data Protection

"Electric circuits are beginning to talk", this was one of the titles in the NZZ newspaper at the 21. April 2010. The data protection commissioner Dr. Baeriswyl is claiming for three properties:

- Data transmission has to be safe
- To spy on the households and to give the data to a third party has to be avoided
- Is it really necessary to collect the data every 15 minutes, this allows someone to get really exact habits

Sensitive information's that are used out of the electric company have to be handled very confidential and it's forbidden to use them for other activities. Providers have to inform the consumers, which are provided by them, at least once in a year, about [10]:

- the origin of the energy
- about the rates for the delivery of the current that are used for the energy carriers

8.4 Providers View of Smart Meter

8.4.1 Introduction

On the energy market the providers can be divided into three groups:

- Producers
- Transporter
- Distributors

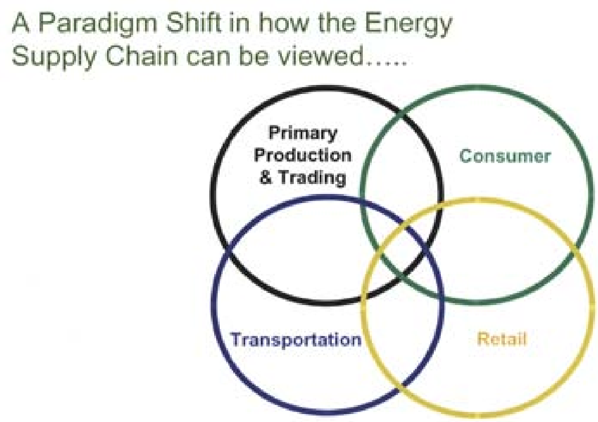

Figure 8.4: Paradigm shift [11]

Most of the company's though operate in more than one of those fields for example a small hydro power plant that also maintains a local low voltage grid and directly sells its energy to the end users. In the classic view, the energy supply chain is organized hierarchical and energy only flows from one direction to the other: From the producers over the transporters to the distributors and finally to the customers.

But the market as it is today does not fit in this schema any more. We will show some reasons for this changes and what role smart metering plays in it. We will also discuss the advantages and problems of smart metering from the view point of a provider.

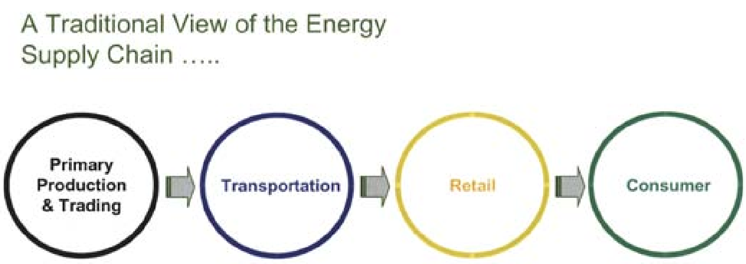

Figure 8.5: Traditional energy supply chain [11]

8.4.2 Market Changes and Driving Forces of Smart Metering

As a result of the aroused environmental awareness of the population the politic demands more and more CO2 saving and energy efficient ways to produce and distribute energy. This leaded to the spread of energy production through renewable resources such as sunlight, wind, water, biomass and others. As the efficient energy production through renewable resources relays on the local availability of these resources, it is often performed on small scale. As a result, there are more and more micro power plants all over the energy net. This kind of energy production is often referred as distributed generation (DG) [12]. DG is an issue for the providers, as the classic view does not work anymore. Energy can now be produced everywhere in the net. And a node of this net can now switch its role from customer to provider and vice versa at any point of time. Imagine a household with a photovoltaic plant on its roof. At one point it might produce more energy than the household consumes, so it injects energy into the net. At the next moment someone switches on a vacuum cleaner and suddenly they consume energy. This generates several problems for the providers. First of all, the energy net is getting more complex. What was a simple one way energy flow in the past, is now duplex. What was once a static net, is now dynamic, as nodes change from provider to customers and vice versa. Another problem is that, the amount of generate energy by private DG is not known to the providers. But to be able to provide a stable net, they need to know. Let assume we have again a household with a photovoltaic plant and it produces about 4 kW at the moment. The household has a total consumption of 5 kW. The local energy provider only sees that the household consumes 1 kW. That's not a big problem as long as it's only a few customers. But what happens if there are 1000 of such plants and it's getting cloudy all of a sudden. Due to these changes, better ways of monitoring is needed. Ideally it should be possible to monitor every node in the net on real time. A task that the old meters are not able to fulfill. They have to be read on side. Something that's normally done once a year or every six month. Smart Meters with their automated data transmission on a fixed time schedule are well suited for this job. Another driving reason for Smart Meters is the will of the politic to give the customers a more transparent energy market. The target is it, to make it easier for customers to judge their own consumption and how much money they have to pay for. With the old meters that's not possible. Only if a customer would write down the numbers every day he could calculate the consumptions of a day or a week, but still not for each hour of a day. As smart Meters can provide that info, some country have even issued laws that force provider to only install smart meters in the future or even replace all old meters.

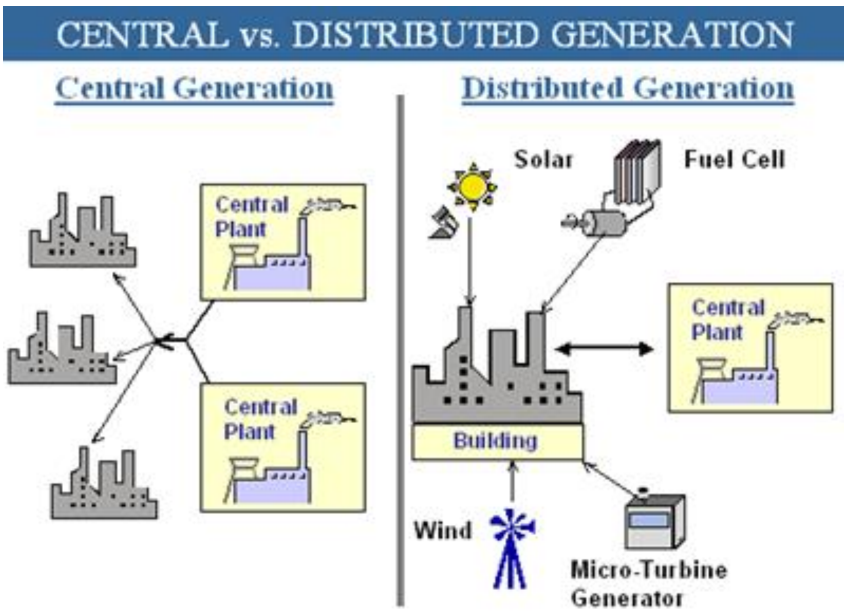

Figure 8.6: Central vs. DG [13]

8.4.3 Problems and Advantages of Smart Meters

There are several problems and advantages tied to the introduction of smart meters. The main problems are the efforts that have to be taken both in terms of financial and personal resources. Another problem is the new technique and the infrastructure that is needed the transport and the processing of the smart meter data. The advantages are (to mention only a few):

- the automated meter reading (AMR)
- the better and faster billing process
- improves fault management
- possible theft detection
- new tariff schemes
- other improved customer services that are made possible through the newly gained information

8.4.4 Problems

For the Installation of the smart meters, there are basically two ways to go. Either you only replace old meters at the end of their live span or you replace them before. Both ways have their own disadvantages. When exchanging the old meters at the end of their life spawn, they are already fully expensed and they would need replacement anyway. But the old meters have life span of 25 years and more. It would take at least one decade till a significant part of the buildings are equipped with the new smart meters. If the old meters get exchanged earlier, they lose money because they "through away" still working equipment, that is not yet fully expensed. Also they have to put in extra work to replace them out of schedule. A general problem lies in the possibilities of the smart meters themselves. The old meters only had to record the total amount of consumed energy. The companies could just choose the cheapest model, which does this job. Different models of smart meters offer different amount of functions. And the cheapest model might in the end be a worse decision. That's something the providers did not need to think about before. And at last but not least the new technical needs. Until now, it was enough to input one number (the consumption of the last year) into the billing system, to bill a customer. Now you have to set up a communication network, which ensures retrieving of the data from all meters at a given period of time. The data has to be stored somewhere in a database, so that at the end of the billing period the data can be retrieved again and the needed number can be calculated.

8.4.5 Advantages

8.4.5.1 ARM

The most obvious advantage is clearly the possibility of AMR. The data needed is delivered fully automatic. There is no need to set up a time with the residents of a house or flat nor to drive there and collect the data. An important note here is that this advantage alone does not out weight the problems the rollout of smart meters causes.

8.4.5.2 Billing

The second advantage for providers emerge directly from the possibilities of AMR. It allows a faster and more accurate billing process. With the old method, the technician had to write down the consumption on a sheet of paper and later copy it to the system. These two sources of human errors are eliminated with AMR. Further it's now possible to bill customers on a shorter period of time without increase in cost, which leads to another benefit. Through the shorter billing periods the cash flow of the company is increased.

8.4.5.3 Monitoring

Smart meters allow the distributors to better monitor their low voltage energy nets. With the information gathered from the smart meters it is possible to real time monitor every node and asked energy usage by DG can be measured. This opens up new possibilities in many areas like load balancing, fault detectionprevention, energy flow optimization, theft detection and others.

8.4.5.4 Load Balancing

Low voltage energy nets do often consist of more than one phase (normally 3). The customers are connected to one of these phases through a relay on their smart meter. The idea is to use more than one relay and connect each relay to another phase of the low voltage net [14]. Now the distributor can not only monitor the load on all the phases, it can also balance it by switching on and off relays on the customer site. For example: Given a 3 phase low voltage net. Let's assume we have the situation as followed: 50% load on phase 1, 30% load on phase 2 and 20% load on phase 3. The distributors now have to choose some customers currently using phase 1 and send their smart meters the command to turn off relay 1 and turn on relay 3. Basically one just switches customers from the highest load phase to the lowest until it is balanced again.

8.4.5.5 Outage Management

There are several scenarios in which smart meters give a benefit for outage management. First we will have a look at the single light off scenario. It's been stated, that about 70% of all outages affect only one customer and that in over 30% of this cases it is a customer problem that does not require a trip to the customer [15]. Let's break this down in numbers. Let's assume that a technician needs at least one hour for the trip to the customer, check the situation and return again (Plus additional time to fix it, if it is really a problem to be fixed on side). Due to the different techniques possible in the smart meters and the communications it might take up to 10 minutes to ping the smart meter and retrieve the information. So on average we save 0.3 * 50 minutes and lose 0.7 * 10 minutes. Which leads to an average time save of 8 minutes. Consider the fact that the most people sleep about eight hours a day and work about eight hours a day. That gives a rather large window of up to two third of the day or even more, where a customer would not realize when an outage happens. The customer would only realize it, when he returns to home/wakes up and has not power. That's the moment when he will call the provider and complain about the problem and the whole chain of fixing the problem start. With the presence of smart meters the provider might long before be informed about the problem. Either the smart meters have some kind of battery and transmission the outage message to the provider during the outage or they stop sending data due to the outage. The second option might take significant longer for the provider to realize that something is wrong, but could still be faster than the customers call. From the point of a customer's view it will look like the provider got faster in fixing the problems. That's because the provider can start before the customer calls. He might even fix some problems without the customer even noticing. While the time needed to repair the problem will not shorten, the time from the begin of the outage to is end does. For larger outage the provider needs to know about 15% of the affected customers to precisely pin down the defect device [15]. While they had to wait for enough calls or send out more than one team to different locations before, now they can just ping there smart meters to see which customers are affected. Again a huge save in time. Another benefit is the confirmation of service restoration. Once a problem has been fixed one can just ping all the smart meters of the affected customers the make sure they are all online again. This minimizes the chance of overlooking a problem when several defects happened at the same time.

8.4.5.6 Theft Detection

With the help of smart meters it is also possible to detect energy theft and (at least in theory) even locate the position in the net, where the theft happens. The detection itself is rather simple. Just sum up the energy consumptions shown by all smart meters in an area and compare it to the total consumptions of this area. If there is a difference you found a possible theft. (Rather than a theft it could also be a malfunction of a smart meter or energy lost due to defect equipment.) While the detection is simple the location is a bit more difficult. An option is to use power line carrier communication (PLC). In this scenario every smart meter would need a power line modem. Power line modems use frequencies modulation to bring data signals onto the energy net. The basic idea is to use special signals that make use of physical properties of the net to measure out the whole electricity net. The main problem of this approach is that every node needs to be shut down to measure out the net, as actual energy usage would bring to much disturbance to the signal [16]. By comparing a new measurement with the original measurement one can tell if there has been a new node added to the net, either parallel or in-between of the original nodes.

8.4.6 Conclusion

While the installation costs for the smart meters and the necessary infrastructure are large burden for the providers, the newly gained possibilities can out weight this burden several times. The main benefit from the smart meters is not the AMR, but rather the new productsmethods setting up on those data and the option to remotely change the setup on the customers place.

8.5 Consumers View of Smart Metering

8.5.1 Consumption

Consumers can control their energy consumption, maybe over a internet portal, and in real time the currently consumption can be read out. The energy consumption can be controlled, some of the devices can be turned off, to save electricity and to save money. The bill would be paid every month, like a normal telephone bill.

We need to change the behavior of the consumption, and for that we have to know what kind of usage we have, this is why we need such indicator like a smart meter. Some people don't know how much they are consuming, and for that it has to be visualized to them. To change the behavior it has to be given a feedback about the consumption, smart meter displays could do this, it shows to the consumer what they are spending, what emissions they are producing, some can give information on specific appliances. Another option is for customers to access information over the internet or to print information on bills [17].

8.5.1.1 Why Smart Meters

Consumers have to understand the benefits of smart meters, and what they are being installed for[17]. Smart meters are the next generation of power, water, and gas etc. counter. Smart meters allow a huge range of new tariffs and capability to sell energy back to the network. Precise measurements is one of the benefits, Providers can exactly tell at which time how many energy would be consumed. Consumer can directly control their consumption. There will be communication between supplier and user. Consumers will be better informed and they are in full control with a real-time display of the energy usage.

8.5.2 Power Eater

Somebody who uses a smart meter is Dr. Germar Büngener, he has installed a smart meter at work and also at home. Every 15 minutes he is watching over the Internet, how his actual power is used when he is turning on and off of some electrical devices. He found a lot of some power eaters. He was really shocked about the usage of the coffee machine in stand-by mode. Today, in the office he turns his hi-fi system completely off. At the lunchtime he is shutting down the seven computers down and saves about 1200 watt per hour [2].

At the moment, there exist two possible solutions, in order that the end user gets comfortable and fast information's, about the actual power usage. LCD-Displays or Internet-based, pc-supported solutions. By the Internet-based solution the assumption is that the user gives personal information about the usage, he has to make a user account and to sign in with his personal id [3].

With the help of those information's it can be shown to the user, which devices use most energy. And maybe he isn't using them at the moment, but they are connected at the grid. These devices are using pointless energy.

Smart meter allow a continuous recording of the power usage and information about the energy that is used. The consumption data can be transmitted and after it can be analyzed. This allows a continuous verification about the power that is used and the energy costs. Whit its help, it can be made a specified analysis about the potential of an efficient energy usage and about the hidden power eaters.

8.5.3 Bill – Financial Incentive to Save Power

In the future, smart meters will enable new pricing options with different rates for different times of the day, week or year. Today, a household receives every year a bill of the cost about the usage. With the adaption of a smart meter, customers will receive every month a bill of the energy usage. To motivate a household to save energy or to motivate to use the energy at other point of time, this succeeds alone with the monthly bill of cost, just partially. The monthly bill is just for the consumer that he can see how much energy he is consuming and that he can choose a better tariff or that he can reduce his consumption through a suitable method [3].

8.5.4 Advantages

One of the advantages for users of a smart meter is the transparency that the user gains about his energy consumption. This knowledge that the user has, can be a great motivation to analyze his behavior and out of that, he can reduce his power consumption. To reduce the consumption and to save money can be a great motivation.

- Energy consumption of the household can be displayed

- Consumers can be informed remotely (historical data) or locally (real-time data)
- Current energy costs
- Customers can control their consumption
- They can switch from a provider to another
- Actual tariff can be displayed
- They can see which provider is the cheapest at the time their usage is at the maximum
- Monthly bills
- Related carbon emission data
- Multi tariff functions can be added to allow demand response techniques
- Allowing electrical appliances to be automatically controlled
- Allowing the consumer to reduce costs by increasing energy consumption during off-peak cheaper tariff periods

8.5.5 Drawbacks

One of the biggest drawbacks is the data protection and the law doesn't really regulate how the data is going to be handled. At the moment, the law requests just the basis version of a smart meter without any options for interaction. It is questionable if the transmission of only the usage data will suffice to reduce the consumption. The data protection is the Achilles'heel of the smart meter. Another drawback of such a smart meter is the privacy that is going to be lost. If it isn't really safe, others can see at what time a household has the most usage and out of this it can be deviated at what time somebody is at home at work or vacations. It can be a nice invitation for burglar.

- The smart meter uses also energy
- Short price explosions because off speculations or a high overall demand of the provider
- Fear of a hacker attack, how safe is the stored data
- Higher acquirement of the smart meter and infrastructure
- Privacy protection is not guaranteed
- Smart meter don't save money and energy
- Smart meter alone is not enough

8.5.6 Indirect Feedback

Energy reduction through indirect feedback is an additional and a very useful arrangement. A very effective instrument is the internet that shows to the user the historical household usage. E-Mail or SMS services that show the actual monthly cost are very efficient, maybe it can also receive a better recommendation. The monthly bill can make a better access to another feedback and information sources. Costumers are very positive exalted about such indirect feedbacks [18].

- Information next day
- Validated for billing
- processed by utility etc.
- mediated through another channel (website, TV, cell phone, billing)
- value-added in combination with historic or comparative analysis for benchmarking etc.

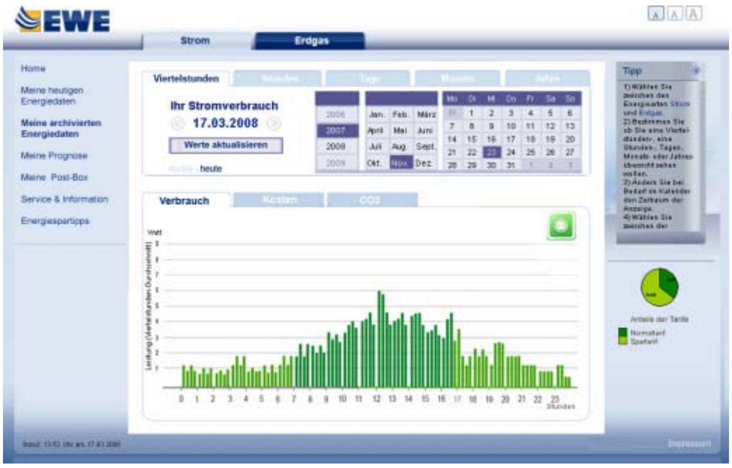

Figure 8.7: Website

8.5.7 Direct Feedback

Display devices that show the usage in real time help customers to detect leaks. They receive a direct feedback of the actual usage. Direct feedback is a very useful instrument to save energy. Households achieve through an inhome-Display savings of about 5% till 15%. The reason for such a high reduction of energy is that the user see's the energy usage of the inefficient devices. The first study in England shows that a lot of the households are not really very motivated to install such inhome-Displays [18].

Another direct feedback can be a mobile end device such as an iPhone or something equal. In general we speak of end devices that are getting information's from content over the GPRS, UMTS or HSDPAD technologies. The design of the information on such end

Figure 8.8: In-home display

devices will be in form of a WEB/WAP sites or text based messages. A feedback over a TV screen shows a special solution, where the costumer gets his actual information of the energy usage direct over a TV channel. There exists a prepayment, where the energy costs can be cleared over an electronic account and this account can't be overdrawn. The costumer gets a strict usage and issue control. These are just a few examples of direct feedbacks.

- Instantaneous and continuous information
- In-home displays
- Not processed
- Present: independent of (conventional) meter: clipon, simple, not validated for billing, not value added

Figure 8.9: Mobile end device

8.5.8 Result from the View of a Consumer

The smart meter alone doesn't give any advantages to the consumer, what the consumers really want, are little helps to reduce the energy consumption. The households are not willing to pay for a smart meter. Not the consumer but the competition will give a chance that a smart meter will be established in a household.

8.5.8.1 Consumer Today

- Unconscious, natural use
- Electricity is just a resource that we can lead the life in the usual way
- Today it is a low control of the consumption and there is no feedback about the usage. The only exception is the bill at the end of the year

8.5.8.2 Consumer in the Future

- Regular control of the own behavior and the usage
- Exact informations of the own usage and the costs
- Try to reduce the consumption, energy is not a natural thing
- Devices with lower consumption will become more interesting

8.6 Summary

Smart meter help to control the energy in a household, but it doesn't reduce the consumption. It shows only the usage and with its help, sources which are using plenty of energy can be found. The energy can't be reduced with just installing a smart meter, the user has be willing to change his behavior and to reduce the energy in the household and for that, smart meter is a helpful tool.

The energy consumption is still largely invisible to the users. People need a feedback and without any feedback it is impossible to change the behavior and to reduce the energy consumption.

Feedbacks can be given in a direct or indirect form, it shows that costumers are very positive exalted about the indirect feedbacks. The direct feedback is a very useful instrument, but a first study has shown that a lot of the households are not really motivated to install such inhome-Displays.

As mentioned before, one of the biggest drawbacks is the data protection, the privacy is not really guaranteed. If the stored data isn't really safe, the privacy is going to be lost.

The law doesn't really regulate how the data is going to be handled, at the moment, the law requests just the basis version of a smart meter without any options for interaction.

The introduction of the smart meter is just at the initial state, some pilot project have to be first carried out to be able to give exact information's about the usage and the behavior of a household.

The future of smart meters will depend heavily on the policy of the government. Today many countries in the EU and outside are involved in projects with smart meters. According to the EU, every household should be vested with a smart meter.

Bibliography

[1] S. Behrendt: *Integrierts Roadmapping, Nachhaltigkeitsorientierung in Innovationsprozessen des Pervasive Computing*; Springer-Verlag, Berlin, Heidelberg, 2010, p. 71.

[2] C. Köhler-Schute (Hrsg.): *Smart Metering, Technologische, wirtschaftliche und juristische Aspekte des Smart Metering*; Berlin 2009, September, pp. 94-95.

[3] C. Köhler-Schute (Hrsg.): *Smart Metering, Technologische, wirtschaftliche und juristische Aspekte des Smart Metering*; Berlin 2009, September, pp. 117-118.

[4] *Smart Metering für die Schweiz - Potenziale, Erfolgsfaktoren und Massnahmen für die Steigerung der Energieeffizienz*; Eidgenössisches Departement für Umwelt, Verkehr, Energie und Kommunikation UVEK, Bundesamt für Energie BFE, November 2009.

[5] G. Deconinck, B. Decroix: *Smart Metering Tariff Schemes Combined with Distributed Energy Resources*; CRIS, Linköping, Sweden, 2009.

[6] Faouzi Derbel: *Trends in Smart Metering*; 2009 6th International Multi-Conference on Systems, Signals and Devices, 2009.

[7] *Netherlands Technical Agreement NTA 8130, Minimum set of functions for metering of electricity, gas and thermal energy for domestic customers*; August 2007.

[8] H.P. Siderius, A. Dijkstra: *Smart Metering for Households: Cost and Benefits for the Netherlands*; `http://www.saena.de/media/files/Upload/smart_metering/PDF/Smart_Metering_NL.pdf`.

[9] S. Darby: *The effectivness of Feedback on Energy consumption*; Environmental Change Institute Universysty of Oxford, April 2006, `http://www.eci.ox.ac.uk/research/energy/downloads/smart-metering-report.pdf`.

[10] R. Dettli, D. Philippen, K. Reinhardt, H. Schäffler, C. Heinemann: *Smart Metering für die Schweiz - Potenziale, Erfolgsfaktoren und Massnahmen für die Steigerung der Energieeffizienz*; Eidgenössisches Departement für Umwelt, Verkehr, Energie und Kommunikation UVEK, Bundesamt für Energie BFE, Novemeber 2009, `http://www.bfe.admin.ch/forschungelektrizitaet/01740/01748/01751/02313/index.html?dossier_id=04264\&lang=de`.

[11] F. MacDonald: *Smart metering: The importance of creating a shared vision - start with a clear view of the end game*; 2009, Smart Metering - Making It Happen, 2009 IET , pp. 1-15.

[12] S. Bruno, S. Lamonaca, M. La Scala, G. Rotondo, U. Stecchi: *Load control through smart-metering on distribution networks*; PowerTech, 2009 IEEE Bucharest, July 2009, pp. 1-8.

[13] Fuel cell applications & issues: `http://www.nfcrc.uci.edu/2/FUEL_CELL_INFORMATION/FCexplained/stationary.aspx`.

[14] A. Pasdar, S. Mirzakuchaki: *Three phase power line balancing based on smart energy meters*; EUROCON 2009, 2009, pp. 1876-1878.

[15] H. Tram: *Technical and operation considerations in using Smart Metering for outage management*; 2008, T&D, IEEEPES, pp. 1-3.

[16] A. Pasdar, S. Mirzakuchaki: *A Solution to Remote Detecting of Illegal Electricity Usage Based on Smart Metering*; SOFA 2007, 2007, pp. 163-167.

[17] G. Owen: *Consumer impacts of smart meters, Smart Metering - Making It Happen*; 2009 IET, Februay 2009, pp. 11-15.

[18] T. Friedli, N.O. Walti (Hrsg.): *Managementguide für Schweizer Energieversorgungsunternehmen (EVU)*; Institut für Technologiemanagement der Universität St. Gallen, Bern, Stuttgart, Wien, 2010, pp. 254-255.