

Internet Economics II

BURKHARD STILLER
OLIVER BRAUN
ARND HEURSCH
PETER RACZ
(Hrsg.)

Institut für Informationstechnische Systeme, IIS

Bericht Nr. 2003-01
June 2003

Universität der Bundeswehr München

Fakultät für

INFORMATIK

Werner-Heisenberg-Weg 39 • D-85577 Neubiberg



Introduction

The Information System Laboratory (Institut für Informationstechnische Systeme, IIS) of the Department of Computer Science, University of the Federal Armed Forces Munich, Germany started research and teaching during this spring term 2002 (FT02) in the area of communications. One of the closely related topics is addressing the use and application of technology and mechanisms under economic and technical optimization measures. Therefore, during the winter term 2003 (WT 2003) a second instance of the Internet Economic seminar has been prepared and students as well as supervisors worked on this topic.

While Internet Economics has been established as an important research and teaching area in the center of technology and economics on networked environments, the underlying communication technology applied for the Internet and the way electronic business transactions are performed on the network have changed during the last ten years. 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

The second edition of the seminar “Internet Economics II” deals with the use of Internet technology and additional ways to support and do business. Starting the talks, a view onto an open market for access networks has been prepared, which addresses technology-wise the competing alternatives and a mobility view on users. In addition, the current state of current mobile commerce sector is addressed in the second talk on its analysis and outlook. The third presentation investigated peer-to-peer systems and started a comparison with a more centralized approach, enlisting a discussion on the economics of such networking scenarios and their respective use. Furthermore, an application-driven view point of Internet communications has been worked out with respect to the use of Voice-over-IP and the current approaches of harmonization efforts between the H.323 and the SIP protocols as well as their respective standards. Finally, the content charging talk investigated the potentials on content offered in the Internet and addressed important mechanisms, which enable various forms of the payment for such content in an electronic fashion.

Seminar Operation

As done beforehand, all interested students worked on an initially offered set of papers and book chapters, relating to the topic titles as presented in the Table of Content below. They prepared a written essay as a focussed presentation, an evaluation, and a summary of those topics of interest. Each of these essays is included in this technical report and allows for an overview on important details, areas of concern, business models in operation, and problems encountered. In addition, every student prepared a slide presentation of approximately 45 minutes to present his findings and summaries to a varying audience of students attending the seminar always as well as other interested students and research assistants. Following a general question and answer part, a student-lead discussion debated lively open issues and critical statements with the audience.

Local IIS support for preparing talks, reports, and their preparation by students had been granted by Peter Racz, Oliver Braun, and Burkhard Stiller. Many bilateral pre-presentation discussions have provided valuable insights in the emergingly moving field of Internet Economics, both for students and supervisors. Many thanks to all people contributing to the success of this event, which has happened in a small group of highly motivated and technically qualified people.

This seminar has proven that the effort to be put into place from students, supervisors, and the professor are significant in the sense, that addressing a good and satisfactory result requires time and reading. However, those who have succeeded and who have believed that this effort is worthwhile, have achieved a phantastic result in terms of presentations, reports (as included in this Departement of Computer Science Report), and final marks. The presenter's insights into a highly relevant area of research, technology, and speculation - surely addressing the Internet environment - have contributed to advances of understanding many problems. Always highly appreciated, discussions after each of the presentation inter-related many aspects of different talks.

Neubiberg, May 2002

Inhaltsverzeichnis

1	An Open Market for Access Networks	7
	<i>Sönke Brecht</i>	
2	Analysis of and Outlook for the Mobile Commerce Sector	31
	<i>Uwe Gärtner</i>	
3	Peer-to-Peer: Die Betrachtung der Wirtschaftlichkeit verteilter Systeme	51
	<i>Florian Gonsior</i>	
4	VoIP - Harmonization of Protocols and Standards	75
	<i>Michael Beise</i>	
5	Content Charging in the Internet	103
	<i>Danny Heerlein</i>	

Kapitel 1

An Open Market for Access Networks

Sönke Brecht

This work introduces the idea of Access Networks (AN) to the reader, names available technologies for having access to Access Networks, and discusses possible markets.

The wish of many users to have access to i.e. the Internet or personal data from any place in the world especially during travelling through a town or country, has made mobility to the centroid of all aspects in this work. One of the main themes is, to explain how static access networks can offer suitable functionality and protocols to mobile users. A brief introduction on a set of technologies for static networks with a dynamic behavior is given. Following this introduction different use cases will be discussed. In addition hints are presented to which extent technologies can be applied to solve the variety of use case-specific requirements. The last part introduces markets, including the problem of accounting, billing and where installing an AN for runtime-mobile clients.

Contents

1.1	Introduction	9
1.2	Access Network - The idea	9
1.3	Technical Issues of Access Networks	11
1.3.1	Mobile End-Users vs. Static Networks	12
1.3.2	Implementation of Dynamic Network Techniques	13
1.4	Use Cases	19
1.4.1	End-User-System in two Separated Networks	21
1.4.2	Moving User in a Company Network	21
1.4.3	End-User-System in Multiple Static Networks	21
1.4.4	Moving Network within one WAN	22
1.4.5	Moving Network within multiple WANs	22
1.4.6	Examples	22
1.5	Accounting	23
1.6	Billing	25
1.7	Summary and Conclusion	25

1.1 Introduction

In the 1960's the Arpa-Net was founded by Universities and the U.S.- Department of Defense (DoD). Basic idea was, to create a network able to hold up communication without depending on a part of the network. This should prevent the military for loosing power over their troops after a possible i.e. a military strike from an enemy. In time of peace the network expanded more and more, departments, banks and companies became part of it and the first private users get access to the Arpa-Net through the telephone network by modems. In this day the Arpa-Net was renamed to Internet and the first companies started to place special Internet-Access infrastructure at the end of telephone lines.

Today having access to Internet is as easy as wearing a shirt. The most concerning of Internet-Access-Companies is to widen bandwidth and trying to earn more money ever before. A second aspect is, today we have a whole industry for Internet. Starting at the hardware level, leaving at the software to being able to promote my person or company professionally in the Internet. Nowadays it is as important to have an Internet homepage as distributing business cards.

But not longer only advertisement is the information one is able to get in the Internet. The proposals are unmanageable. Nearly everything is accessible. Scientific papers, latest news, personal mail and street-routing is also available as the newest rumors mentioned at the Internet blackboard called news-system.

The following questions remain unsolved: How is it possible to connect? Which are the techniques? Is it only a static network or is it possible to walk around and having access like mobile-telephony? What are the costs for this luxury?

Due to the fact, that immobile solutions are far more developed than mobile ones and the market for the static network access is in the hands of global players like German Telecom, Swisscom, France Telecom, Bell and others I am going to discuss theses questions with a mobile background.

In the first chapter the idea of ANs is introduced. Differences between a client, an access network, and a core networked are explained. The second chapter will inform about mobile access solutions. Chapter three will show possible markets, chapter four is having a look on accounting, chapter five will ask how to bill and chapter six will give a conclusion and show a possible future of ANs for runtime-mobile clients.

1.2 Access Network - The idea

The first question arising when the term Access Network is used, is what is meant by this expression. Typically there is no global definition for Access Network. The ITU Study Group 15 has made a valid and applicable definition of AN: „The access networks (AN) provides connection between customer's premise Equipment (CPE) and a set of service nodes (SNs), either directly connected or remotely connected via a transport network.“ [1].

According to this, a network, interconnected to any other network, is an access network for the members of this network to the foreign network and vice versa. Interesting is the point,

in this definition is not said who gives access to what, too. In this way access and network are very abstract. In this paper these two terms are reduced to to the aspects of computer networks and having access to these networks with runtime-mobile network-clients. To give examples from whom an access network can be used have a closer look to [Fig. 1.1] Three types of access-network-clients are shown: Mobile phone and phone, computer and

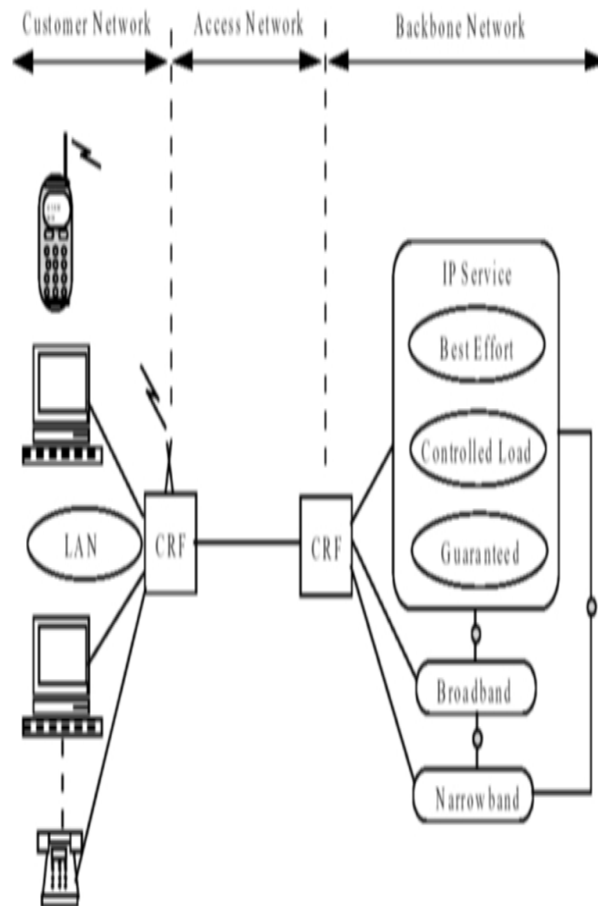


Figure 1.1: Common access structure [1]

computer over phone-line, and an entire LAN (Local Area Network [2]). In the middle of the picture there are two CRF (Connectivity Related Functions)-points at the two ends of an AN and on the right a Backbone Network is drafted. The Backbone contains everything users want to have access to. The AN provides access to this backbone. In this context the AN may be a physical or logical connection to the Backbone¹. For example a AAA (AAA = Authentication, Authorization and Accounting) server can be the logical connection to a corporate Network or email system. The telephone for instance needs a telephone network, the LAN needs a computer network e.g. a x-DSL (x = Type of DSL; DSL = Digital Subscriber Line). Therefore, AN is an abstract concept, implementation depends on the service(s) the clients wish to have and the backbone offers. Due to the fact that telephone networks² are very common and this technique compared to i.e WLAN (Wireless LAN [2]) is quite old, does not offer a new market (in case of offering access, technique

¹Backbone in this graphic also includes core networks

²meant are static and mobile telephone networks

itself might be better in future), the focus is laid on computer networks providing users or clients access to them while these are moving.

According to the discussion above key questions of this seminar work are:

- What type of mobility is needed and how is it to be implemented?
- Which kind of access a user can have?
- Which technical solutions exist?
- What has have to be developed to add (more) mobility to existing solutions?
- Who pays for what?

Mobility is the key aspect of this seminar work. Having a look at the development in telephone networks views: 20 years ago the first mobile phone networks started and the days of the phone boxes were counted. Today both are co-existing but phone boxes are in reduced presence. At places where are a lot of people like public places, shopping areas, train stations and so forth the first mobile connectivity points were installed. With the time the mobile-phone networks were expanded to support mobile telephony everywhere. With the expansion of the mobile-telephone network phone boxes were abolished, except former named places.

A few years ago the first wireless networks were installed at corporates. Today the first mobile-computer networks for public places are available, sponsored by public administration. This was made possible by reducing the size of high performance computer from mainframe-size to a little PDA. Another evolution was necessary, too: From inventing the Arpa-Net to WLAN. On the one hand there are clients, easy to carry and moveable, on the other hand physical network connections do not depend any longer on cables.

In summary, the goal is determined by the fact to achieve a higher degree of mobility with similar functionality and a higher integration of services

1.3 Technical Issues of Access Networks

Caused by technological developments since the invention of the first telephone and inventing wireless Morsecode transmitting until today spotlight of this chapter is pointed on technologies available and required for mobile access-networks. Basic concept for works about network technologies is the ISO/OSI Base Reference Model (BRM = Basic Reference Model[5] shown in [Fig. 1.2]):

Considering the model some questions are, which Layers are directly impacted by mobile networks, above which layer is it like the same behavior for higher layers like in static networks? What has to be done making available movements during runtime of a client or giving networks dynamic features? How far are dynamic networks implementable and which solutions or implementations do exist. Most of these questions will be answered within the next chapters.

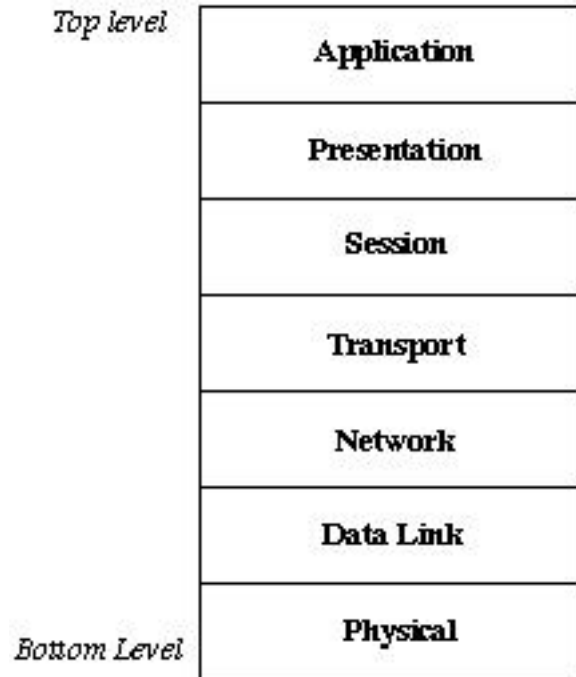


Figure 1.2: Common access structure [3]

1.3.1 Mobile End-Users vs. Static Networks

Like everywhere in the computer technology there is a great distance between user or customer wishes and state of the art technology. On the one hand it is a disadvantage, because customers are unsatisfied. On the other hand this difference is the origin of technological processing. The origin of mobile networking is the wish of the user to have network access i.e. during moving through town, sitting in an airplane and so on. Registering these examples in a more abstract way, a so called user wants:

- Ability to move in room at runtime
- Connectivity everywhere
- Network behavior like mobile telephony
- Uninterrupted access to personal data (Email, news, ...)

As the counterpart static networks provides:

- Access to personal data (via: ssh, ftp, pop3, nntp, ...)
- connectivity where a network-port is (e.g.: POTS (Plain Old Telephone System))

Summarizing these two lists, mobile network technology has only to ensure the connectivity at runtime to a static network. If we were able to build a bridge between exiting networks and moving clients, we can use the static networks as backbone or core network.

Remembering the scheme [Fig. 1.1] we have now a well defined situation and can start solve the problem. An AN is the technology to connect mobile clients to the backbone network.

In the first list two outstanding claims: The ability too move around in room and noninterrupted access to the core networks. These claims were the same like in mobile telephony as it started to become public. To allow free moving in a room we need a technology independent from wires between the mobile client and the core network. Today's solutions are implemented with short range radio technology, like mobile telephones. The other point, having uninterrupted access to the core network is a bit more difficult to solve, even if you think of long range and high speed movings like travelling in an airplane or train.

Now major technologies providing dynamic access to a network are introduced.

1.3.2 Implementation of Dynamic Network Techniques

The two claims, the ability too move around in room and noninterrupted access to the core networks condition two different approaches. The ability to move free without having to look for wires requires a wire independent physical layer solution. Noninterrupted access to the core network independent from the place the user is requires a solution at the session layer or above. The layers between have to be remanaged, too.

Networked Desktop

Mobile is available telephony for 20 years and Wireless LAN is for a few years available. The time before millions of telephone-boxes respectively computer-terminals were used to emulate some everywhere connectivity. Basic Idea of this concept is, a user is sitting at his desk, working at his³ workstation or terminal with a user account. Later he is changing room to anywhere and there he wants to continue his work⁴.

The solution contains two major components:

- A central server
- A set of workstations or terminals.

The server contains every data related to the user i.e. bookmarks, email, user-account related data, applications (binary and personal preferences), specializes variables for the OS and so on. At the very beginning of computer networks, the calculating power of any terminal was also provided by this or another server. Often there were connected two or more server to one large cluster (i.e.: one for user related data, one for applications, one for computing power).

³In this case his stands for male and female users

⁴E.g.: A technician helping a worker in a special matter by using his administrative privileges

The user itself take a seat in front of terminal or workstation connected to the server⁵ Later, as it was possible to put more and more computing power into a terminal, from now on called workstation, the centralized data and application was transferred into the workstations. With increasing capacity and availability of harddisk-space the central servers only a quotient of data was transferred each session.

The idea behind networked desktop is: To have within a network one central place where the changing and user specific data is stored. This allows users to walk from one workstation to another, having everywhere „his“ desktop after logging in at the terminal. The terminal loads the necessary data from the server. This was the first step, others followed:

BootP and DHCP

BootP (BootP = Bootstrap Protocol) and DHCP (DHCP = Dynamic Host Configuration Protocol).

After having workstations, the laptop was invented. A laptop is a portable computer, not that powerful like a workstation, but it could be carried around with the user. Now the people did not have to look for a free terminal anymore, they just had to find a free network outlet and reconfiguring their computers for the needs of the network they were in.

This solution was not able to satisfy all users. Not everyone was willing to know how to reconfigure the laptop by himself. The computer specialists invented an automatic system, reconfiguring the computer for the network needs. This technique uses the fact, that every ethernet network interface (NIC) card has got its own special number: The MAC (MAC = Medium Access Control)-Address.

According to [Fig. 1.2] the MAC-Address is placed in layer 2. PDUs (Protocol Data Unit) in ethernet networks [4] got a header by the sending network card with its own number and the goal number. These sender number is used to identify the computer connecting to the network by a BootP/DHCP-Server.

BootP is preversion of DHCP. DHCP offers some more features like BootP. BootP is fully supported in the so called static variant of DHCP. Basic idea of both is, a computer is connecting to the network, sending out a request message by broadcasting into the network this request. The DHCP-Server is able to identify the computer by his MAC-Address. After identifying it, the server sends an answer with all necessary information like the IP-Adress⁶, Gateway and other network related information. In [fig.1.3] is this process painted:

Main difference between DHCP[6]⁷ and BootP is the ability to allow not known computers access to the network. This feature was introduced with DHCP and has got two different modes (dynamic and automatic) useable parallel to the static mode.

The static mode is a table, allocating an IP-Address to a MAC-Address. The other two modes has got an pool of IP-Addresses, the allocate to unknown MAC-Addresses. In the

⁵in this environment the terminal is the AN and the servers are the core Network

⁶IP = Internet Protocol. Most common protocol in the world with over 90% use[2]

⁷For a more detailed overview read [8]. Additional definitions options of DHCP are defined in [7]

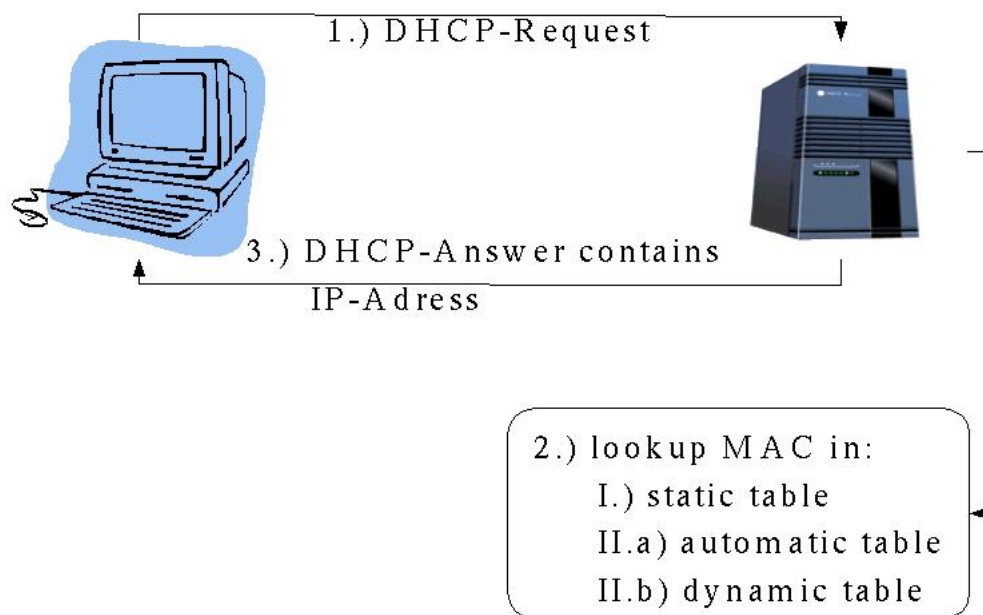


Figure 1.3: DHCP session

dynamic mode a computer got its IP-Address from the pool for a certain time, in the automatic mode this time is not limited.

For mobile clients is BootP not useful if the client want to join a network that does not know him. DHCP is useful for mobile networks if it used in dynamic or automatic mode. There an unknown client is able to connect and use public services of the network.

Mobile IP

Due to the fact that the Internet Protocol (IP[10]) was designed for static networks and the fact that the IP is the protocol used nowadays, the Internet is not able to provide support for mobility. A workaround has to be developed, mapping the mobile behavior of clients to static networks. One approach is mobile IP[9].

One solution is, caused by the static structure of the Internet, changing the IP-Address of the mobile client every time it leaves the range of its current (sub-)network. This makes seamless roaming as known from the mobile telephony infeasible. The change of the IP-Address cuts the connection between the sender and the receiving mobile client. This causes the end-to-end connection to be interrupted until a new connection between the two communicating computers has been established again. The second problem for e.g. Voice over IP is, to reidentify the host with his new IP-Address.

According to this, another solution was invented: The Gateways should, if ordered to do so, track the way of a special signature placed in the IP-Packets of the moving client. If the client changes its IP-Address, the signature would identify it. The first router, who receives this tracking signature from the new IP-Address, would reroute the lost packets.

This solution would cause several changes to the implementations of the Gateways and routers and requires a special date where this system is able to be used by everyone.

Another answer has to be given to the customer for the problems of ad-hoc availability and uninterrupted data-streams. A network structural prerequisite is the change of the IP-Address of the moving client per network he joins. But usually these hosts have got a network where they belong to, their home network. There they have an IP-Address. This means a host has got two IP-Addresses if it is not in its home network belonging to it, his home address and a foreign address. The idea is, to let the communication believe, the mobile client is at its home network. In the home network a special computer, the home agent (HA), knows the actual foreign address of the moving client and readdresses the packets to the moving client being sent to his home-address. For Mobile IP three additional components are needed: The mobile client, supporting mobile IP, a HA, and a foreign agent (FA).

The mobile client, or also called mobile Host (MH), has got to know its home IP-Address and HA.

The HA knows all mobile clients it is responsible for, knows if they are connected to a network in which network they are at the moment, tracks the position of each client with the help of the FAs, and intercepts data streams to mobile clients if necessary and forwards the packets to absent clients. The FA informs the HA of a client about the presence of a specific client in FA's network and allocates a foreign IP-Address to the mobile client. After having done all administrative at the beginning of a session, the session itself works this way:

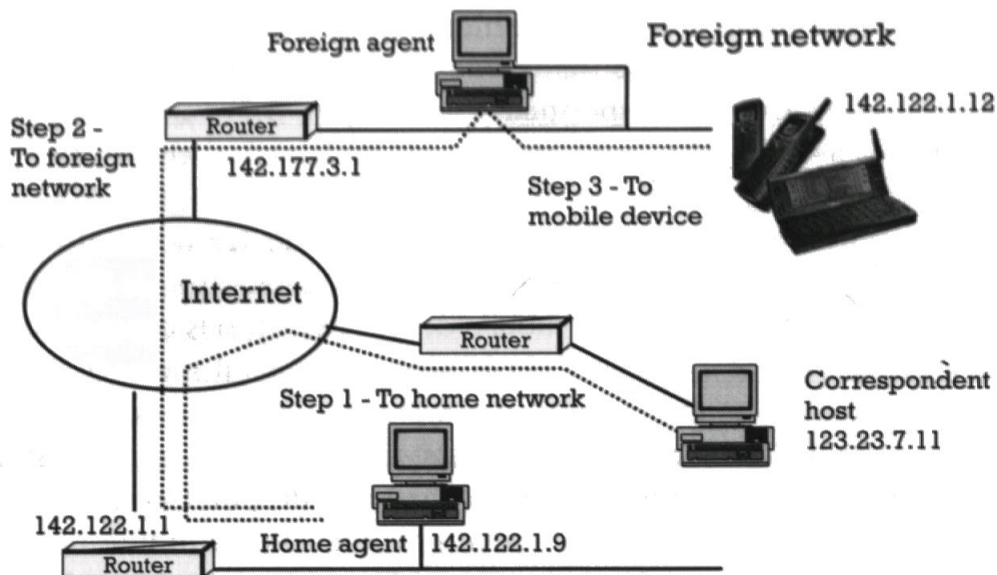


Figure 1.4: Mobile IP communication [11]

In [fig. 1.4] you can see the Internet, a correspondent host, three routers, a FA, a HA, and a mobile device. In step one the correspondent host sends an IP-Packet to the home

IP-Address of the mobile device. The router of the network of the correspondent forwards this packet through the internet to the home network of the mobile device, because the sink address of the packet is the home address of the mobile device. At the home Network the HA knows that the mobile device is absent and where it is. This causes the HA in a second step to receive the packet instead of the absent mobile device and adding a new IP-Header to the packet and readdressing the packet to the foreign network where it is (new sink address is the IP-Address of the FA). The router of the home network routes the packet to the internet and it is routed to the foreign network. The FA receives the packet and removes the second header in third step, sending the packet to the mobile device which finally receives the packet.

Advantages of this method are the possibility to reroute packets to the actual network where the mobile device is in just changing the allocation table of the HA, notified by a FA. End-to-End connection are probably interrupted for a dozen packets for the reallocating process while the mobile device changes its current network, but have not got to be reestablished. If you have a look at the round-trip-times are they no problem to the communication, because the packet has got to run all the time from the correspondent host to the HA and from there to the mobile device. If the mobile device changes its network, some packets are lost and can be resend easily by the correspondent host if necessary. This enables Mobile IP to be used additive to the current gateway, router and other IP based communication architectures.

Mobile IP with DHCP

Here is a special, very common, instance of Mobile IP discussed. In the text above was mentioned the FA would assign an IP-Address from its network to the mobile device. This can also be done by a DHCP-Server. The Advantage of this solution is, if there are mobile device that do not know their home IP-Address or there HA or their home network is unreachable, they can take part in the specific network like any other guest. Behavior of this mobile device is the same as if their were no Mobile IP invented.

Case the mobile device and its home network are ready for Mobile IP, the Care of Address (CA) is also assigned by the DHCP-Server, but all advantages of Mobile IP can be used.

Layer-1 Technologies

There exist a number of solutions for having wireless technology for connecting to a network e.g.:

- Wireless Local Area Network (WLAN)
- i-Mode
- Universal Mobile Telecommunication System (UMTS)
- UMTS Terrestrial Radio Access Network (UTRAN)

- General Radio Packet Service (GPRS)
- Global System for Mobile Communication (GSM)
- Bluetooth
- HiperLAN2

But most of them do not fit all requirements of modern network infrastructure like large bandwidth per connection, high amount of connections, fast signal transmission, stable links, low cost. The criterium largest possible bandwidth per connection is provided by WLAN with up to 22 Mbit/s. UMTS provides 2 Mbit for all users connected to one antenna. Bluetooth either provides a large shared bandwidth. GPRS and GSM both provide not more than 64 kBit/s, too low for today's needs. The i-Mode Standard itself provides a special service for mobile phone sized devices and bandwidth is with a maximum of 9.6 kBit/s not really useful for bringing a Laptop online. However, how much bandwidth a single user desires, depends on the user itself, but just waiting more than 20 s for an email or several minutes for a webpage will not be acceptable for a user.

Two candidates are left: UMTS and WLAN[12]. UMTS has compared to WLAN a very small bandwidth, especially when a lot of users have to share the small bandwidth. Disadvantage of WLAN is the short range, ordered by government. Caused by the rare and already sold⁸ radio frequencies for UMTS, these technique does not interest for this seminar work anymore. The market is fixed and not open. Experiments are done by the great telecommunication companies and not by new probably growing companies, the market is closed.

A special role is given to HiperLAN2 which is the European counterpart to WLAN. It is compatible to other network systems: Bluetooth, WLAN, UMTS, and IP. Caused by the experimental status of this project it is not discussed in here. For further information have a closer look to [13].

WLAN

Wireless LANs were developed to allow short range movings of a WLAN-Host like DECT-telephones for wireless telephony at home or office. Mobile IP enables home network independent access to the home network and WLAN add physical mobility to mobile hosts. The best thing with WLAN is, hardware prices are on the decrease and the mass of customers is willing to buy this technology. A moderate bandwidth, depending on the distance between host and base station, with today up to 22 Mbit/s is ready to transfer even large files within acceptable time.

A structural disadvantage is the decreasing bandwidth if the host moves at the corner of the range of the base station as shown in [Fig. 1.5]. The next disadvantage is, that the range of the base station is limited to 10 to 15 meters. This causes the need of a lot of base stations to provide wireless access at a large place, for a great amount of users or a complex building. Another category of problems are security and privacy related.

⁸In Germany and other major European countries

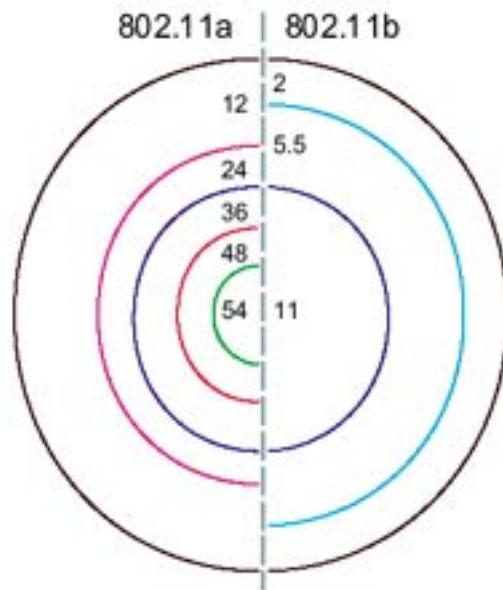


Figure 1.5: Distance dependent Bandwidth [20]

Caused by using all direction spreading radio technology it is very easy to listen to the data transfers. For solving these points there were developed different mechanisms, like masquerading the data signal and/ or randomized changing of transmitting frequency[14].

1.4 Use Cases

Reality is quite complex and schemes has to be identified for better understanding the coherences within reality. In this chapter a classification of different network related use cases will be made and exemplary discussed. In this classification the view of users dominates the classes and decisions and will be a view into practice. Classification will depend on three variables:

- The user that is moving or not
- The client that can be moved at runtime or not
- The count of networks the moves of both are done

These variables can be shown in the following table:

User/Client	not moving	moving in one network	moving in n>2 networks
not moving	any	not a case	not a case
moving in one network	networked Desktop, wall outlet	wireless	not a case
moving in n>2 networks	DHCP w/ w/o Mobile IP, networked Desktop	not a case	wireless w/ DHCP or Mobile IP

In this table are all possible combinations out the three variables presented. The case of a not moving user with a runtime unmoveable client represents the classic static case. Example for this is a user sitting at his desk with a traditional workstation he is working with. A lot of solutions for this case are existing today and the market is quite old. Paying attention to the other two cases in the same line: They are both marked with „not a case“. This is done because the ability of the client to be moved within a room or different networks will not be used by the user at anytime, but and that is important, this ability hurts no one and causes no troubles.

In the second line the user wants to move through the room or within range of one (sub-)network. If he has got a runtime immobile client e.g. a laptop with a traditional wire based ethernet network interface, he has got to pause his work, change the room and search for another wall outlet where he can plug in his laptop and continue his work there. Another example in this case are workstations within a building and the user has got to change the room. This causes him to pause his work, logging of from the current workstation and logging in at another workstation where he can continue his work. The system offering this service is called networked desktop.

If the client has got to be runtime mobile. WLAN can be used for this. In general wireless network connections are required for being runtime mobile within a network. If the ability of moving within different networks at runtime is offered, it will not be used anyway in this scenario, again, this ability want disturb the wishes of the user in the two cases before.

If the user wants to move within multiple networks with a runtime immobile client, only the possibility to continue working by using for e.g.a SSH session. Other possibilities are to connect in another network by using DHCP to reconfigure the client and having access through a layer seven application. Mobile IP can also be used or networked desktop if the foreign terminal provides this service.

The second case in this line is not a case because it is a maybe wireless to network connected client. But its behavior is the same like in the case before. If the home network is left, there is no runtime work possible. This causes the user pause work, too, and reconfigure the client for the new network.

The last case is the very interesting one. Here is network independent runtime mobility mapped. The assumptions of this case is a user who wants to move large distances or not foreseeable ways and has got a runtime mobile client. As shown in the cases before runtime mobile clients needs a wireless connection to a network. But this is not enough, because different networks will be used during work. Here a special solution has got to help, managing the traffic if the client together with the user has changed the network. This solution is provided by the different instances of Mobile IP.

Now it is time to discuss different examples which have been identified as typical scenarios for using a client by a user. It is shown that the ability of the client to be able to move in multiple networks want cause problems.

1.4.1 End-User-System in two Separated Networks

Imagine a Laptop used by an office worker. Some work has to be done at home and he wants to use his client in his home network which is providing access to the Internet where he has to gather some information. After plugging in to his home network the client has got to be reconfigured for the network parameters of the new network. This can be done manually (not that exhilarating), or by a network service. One service for automatic reconfiguration is DHCP. But if he uses Mobile IP, he has got the same behavior of his client like in his office. As you can see, using Mobile IP with the ability to move during work can hurt the user.

Nevertheless this example offers no new open market, because the access network is already installed and in most cases the access to Internet is provided via telephone and/ or DSL.

1.4.2 Moving User in a Company Network

Here a user has got to move through the building. Until he is moving, he wants to upload or download some data that should be available when he arrives at his destination. Two basic cases are possible: Where he moves a workstation is available or not. If, networked desktop can be used. If not the data must be transferred to or from the client he is carrying with him. This can be done by a wireless connection.

Imagine it is a huge building and different subnetworks have got to be crossed. Again some technique is needed to manage the datastreams. Mobile IP or similar techniques can be used.

Unfortunately an open market is not given in this example, because the wireless technique and the Mobile IP agents are additionally installed to the existing core network of the company and network access is free for all employees.

1.4.3 End-User-System in Multiple Static Networks

Walking through town and receiving information about the place where you are, or talking to a partner through the internet while moving from central station to airport. Who has not had this dream? This dream is not a dream any more. With Mobile IP and a wireless network connection this can be done. And here is given an open market for different parts.

The ones who provide access to the internet want to earn money. Serendipitously this kind of access network is not very common yet and the networks existing are provided or sponsored by government. This sponsorship will end one day and there will be a new market for companies.

1.4.4 Moving Network within one WAN

What has to be done when a whole network is moving? A possible scenario is a train moving on a railway through Germany from Hamburg to Munich. Here you have to manage hundreds of clients moving at high speed. But there is an important advantage: The way of the train is known. Network access points for the whole train can be installed on the way. Between train and the point is maybe a wireless connection, between access points and core network maybe wired connections. Due to the fact the way and the time when the train is in optimal range of the next node, a layer one routing of all the data streams can be installed. Within the train the network would behave like any other public access network.

This service maybe free, but after the customers are used to have the train company can earn money with it. Stupidly only the train owner can install the network. No open market has been identified.

1.4.5 Moving Network within multiple WANs

Let us widen the former example to moving through countries or continents. Think of an airplane moving from Germany to Australia. Several hours flight time. Bored passengers, forced to sit around and consuming on board entertainment instead of using the time for checking emails, answer last minute calls, etc.. This wasted worktime can be used by connecting the airplane via satellite with the Internet. Caused by the high costs of installing satellite links the market is open to companies with the know how and enough money. But with a good strategy maybe a start up can also be establish itself there.

1.4.6 Examples

The market open to everyone is described in the third example. Nowadays a lot of free public access networks are installed. But also commercial access networks are installed like in Star Bucks restaurants. Other for commercial access networks maybe places like Airport lounges, railway stations, subway stations, buses, trains, airplanes, parks, and highways. This list of examples is not complete anyway. In general at every place where a lot of people are at the same time with the chance to sit down for a while open access networks can be installed. Some public open AN are:

- WoomS e.V. (Worms)
- Bürgernetz Aichach-Friedberg
- Amsterdam
- Munich Scientific Network (provided by lrz-muenchen.de)

WooMS e.V.

„WooMS e.V. (Wireless open organization Muenster e.V.) [16] is a non-profit registered association since 3 June 2002. It's goal is the structure of an opencitizen net over radio technology. The technology, Wireless LAN after 802.11a/b, offers fast mobile data communication with up to 54 Mbit/s. It is license free and free of charge and thus for everyone freely usable. As a result of Wireless LAN arise many new application types, of which we for the advantage of Muenster, which would like to use promotion of science, education and culture.

This radio technology engineering offers for the first time the possibility to create and operate transnational as well as independent net infrastructures to the citizens. This means for WooMS primarily once the structure of a fast and safe Backbone over the roofs Muenster's, as it were the backbone of the net, secondly the filling with local contents of (project Muecon, a derivative of Muenster's content). Among them we understand e.g. cultural meetings which in high quality live in the radio and TV are inserted and thirdly the installation of Hotspots, mobile net entrances at public places. Thus in the road cafe sit, to surf, write email or on-line play and for in vain;) Further projects are in planning.“⁹

Munich Scientific Network

The faculties of Munichs Universities are allowed to provide network access via wireless technologies within their buildings. A lot of rules and guidelines have got to be met. The user is allowed to take part in the network for his scientific work. The propagation of this network is shown in Fig. [1.6]

1.5 Accounting

A great problem of this commercial open Access Networks is providing access. After commercial providers have got installed the hardware they want to earn money. The user, willing to use the provided service has got to get an account allowing him to take part. In case of a subway net or of the town he lives, places he often visits, where the user moves all the day to his work he is able to register to all of these, maybe for each location different, providers. But in a lot of scenarios and cases it is not possible for the user to register in advance.

Imagine the manager of a large car construction company in Stuttgart, Germany, has got to travel to a overseas branch in Detroit, USA. His bureau will book a flight ticket. Maybe they can not get one from Stuttgart but from Frankfurt. This causes they have to buy a train ticket, too, and a car to the train station is needed. During his travel the manager want to inform himself with actual data about the branch and other to this location related details. In the worst case this cause him to have an AN account for every AN he is going to use at. One account for the AN of the train company, for the airport

⁹Translated from German, source [17]



Accesspoints im Münchner Wissenschaftsnetz

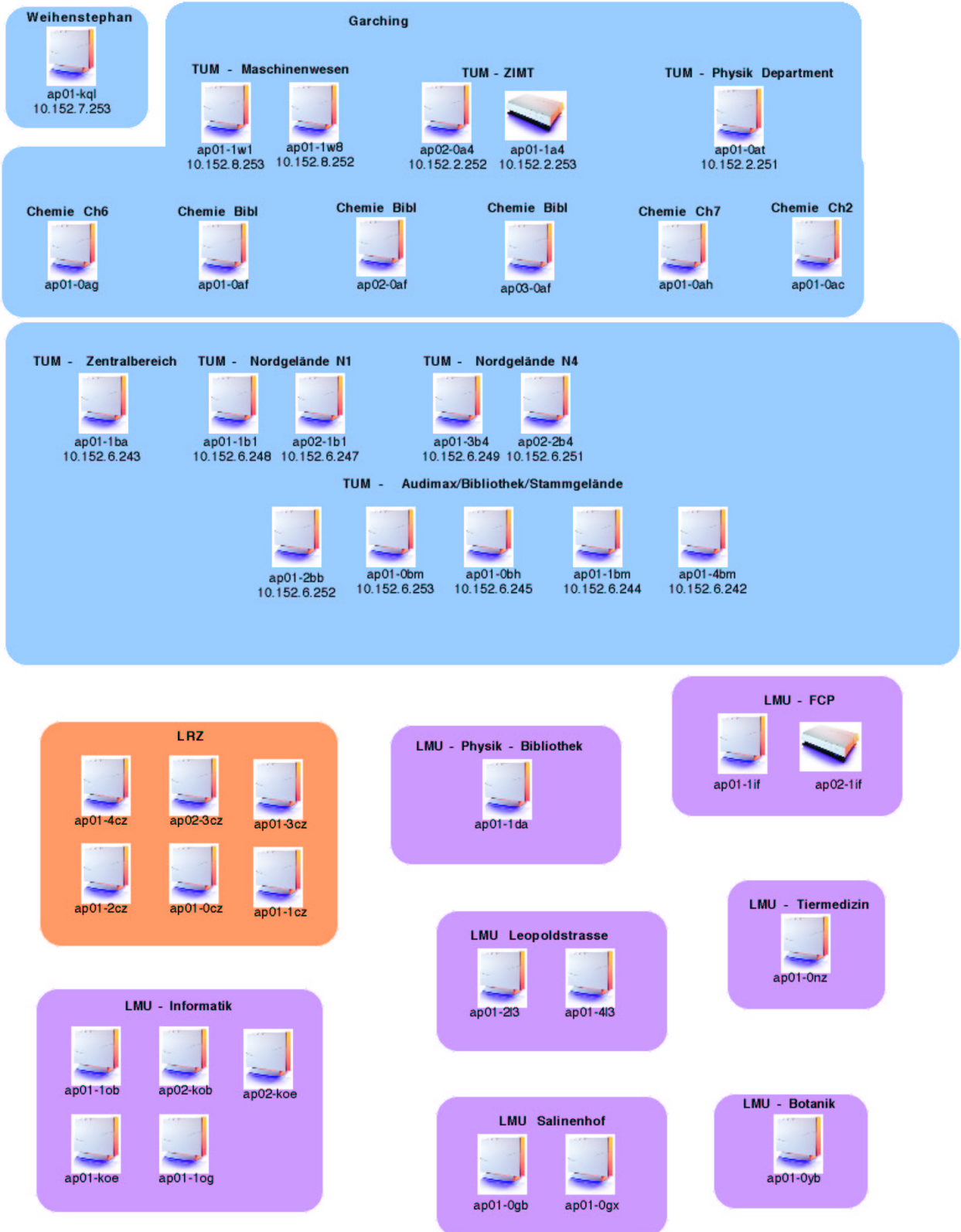


Figure 1.6: Accesspoints within Munich Scientific Network [19]

in Frankfurt, for the plain and for the airport in Detroit, for the town of Detroit and so on. Every change of provider will cause an interruption of his actual session with entering new account data.

All the registration work for the single AN providers can be done in advance, but if it necessary to change route cause of late trains, planes, traffic jam or anything else, it can not be done in advance. Due to the fact, that the information is already desired, some other way of accounting should be invented.

Three possibilities exist: The multipolistic as discussed above, an oligopolistic and an monopolistic solution.

The multipolistic solution is not comfortable and flexible enough for users, as seen above. The Monopolistic solution is at the moment not possible, because of two facts: This solution has got to be an international one. According to the great divorces in international relationships today, there will not be a consent possible. The second point is, who will oversee this non-profit organization? Every country taking part?

Only the oligopoly solution will be the right one, at the moment. But they all, users and providers, have got to trust these accounting services. In [15] this solution is discussed in detail.

1.6 Billing

The installation of commercial open access networks causes the flow of money between user and provider. The user has got to pay for the service he used provided by the provider. Remember the pricing scheme of an ISP (Internet Service Provider). E.g. one minute online costs 1.5 ct. If a person works everyday at the same places and travels every day the same route or within in the range of the networks of the same provider(s), at the end of the month he will have been used the service for several hours and the bill will be about EUR 20. But what will happen, when the user has just used the network for three minutes within a network he will use only once a year or decade?

Shall the provider send a bill for 4.5 ct, by paper, printing, and post costs of ca. 60 ct? If the number of to low bills are only a few it might be acceptable for the provider. But what to do if you are at a place where most of the users are only for a few hours or minutes like airports, lounges, trains? Again in [15] a lot of solutions are discussed. They are not especially for open ANs but solutions for so called micropayment (bills with a worth of less than one ct.) are introduced. These technologies might provide a solution for the losses of providers.

1.7 Summary and Conclusion

The key question of this seminar-work was: “Where are (new) open markets for Access Networks?,,.

This question was reduced to the part, where are access networks needed and which services are needed and how much data has got to be transferred between the network and one client.

Assuming a wireless access to the network and the wish of the user to move following scenarios are possible:

- the user moves himself
- the user is moved

If the user moves himself by e.g. driving a car he certainly wants to be callable all the time, wants to know about traffic jams on his route, his actual position and maybe breaking news. Callability is given by mobile phone networks, traffic information is given by radio and TMC (Traffic Media Control) for routing software, his actual position is presented by GPS and breaking news are delivered by radio or phone.

But what happens if he has got a driver? The driver also needs the information about traffic jams, actual position and other car and car driving related information. But the person sitting in the back seat wants other information. Information that are not measurable by car or broadcastable by radio. So he needs an access network to connect with the internet. For downloading files or gathering information a connection via mobile phone is too slow and far too expensive compared to traditional internet by phone connections. Wireless broadband technologies are UMTS and WLAN. UMTS is no open market, because the frequencies are property of the large telecommunication firms.

The question is, which technique will survive? GPRS, UMTS or WLAN?

There exists no answer today, only hypothesis. Which technique will be cost-effective is decided by the users. The more users will use an AN at a special location the greater is the chance this specific AN will survive. But this causes on the other hand, existing technology pointing in the same market will be wiped out.

A WLAN AN for instance cannot survive in an area with great circumference and only few users. These is caused by the design of the technique. UMTS as wide range broadband technology can be used for this. But UMTS is no open market any more.

WLAN is the only technology left for open markets at the segment of ANs. Boundary conditions of WLAN reduces the field of applications to locations with a large number of users with a place to sit down¹⁰ and work.

Locations meeting these requirements are e.g. trains, airplanes, lounges, airports, cars on highways, public parks, public places, and shopping areas, and public transport.

Open Markets are not possible at locations property of a company. This company will install it's own AN and sell the connection to the users. Maybe they will not do it by themselves, but a special provider, partner of the owner will do this.

¹⁰Peoples are running cannot surf that well

Also at public locations there will be a kind of competition before the installation of the first antenna. This will be the open market. The competition before the network exist and if an offer was to bullish again, when another partner for the possessor has got to be found.

Rivalry between UMTS and WLAN at this places is not given, because these two technologies are not competing at the same clients. UMTS is designed for mobile telephones and WLAN is designed for computer based networks. Interfaces between these two concepts and combined solutions exists, but the combined clients are too expensive for the standard user. The WLAN solution is a solution for users with special purposes where an interface to UMTS maybe will not be cost-effective enough due to the small shared bandwidth compared to WLAN.

Bibliography

- [1] ITU Study Group 15: *Building and Operating Braodband Access Networks*, Project P917, July 2001
- [2] B. Stiller: *Protocols for Multimedia Communications* (Scriptum for lecture), University of Federal Armed Forces Munich, Germany, Information Systems Laboratory (IIS), 2002
- [3] Diagram of ISO BRM Layers, <http://www.cs.cf.ac.uk/Dave/Internet/7layer.gif>, 03/15/2003
- [4] IEEE 802.3: *CSMA/CD Access Method*, <http://standards.ieee.org/getieee802/802.3.html>, 03/15/2003
- [5] ISO/IEC 7498: *Information technology – Open Systems Interconnection – Basic Reference Model*, <http://www.iso.ch>, 03/15/2003
- [6] RFC 2131: *Dynamic Host Configuration Protocol*, <http://rfc.net/rfc2131.html>, 03/15/2003
- [7] RFC 2132: *DHCP Options and BOOTP Vendor Extensions*, <http://rfc.net/rfc2132.html>, 03/15/2003
- [8] Marco Huppert, Stefan Laurs: *Seminararbeit DHCP - HuLa 2001*, http://dnserver.nt.fh-koeln.de/studienarbeiten/ws00_01/1/#2.2.1, 03/15/2003
- [9] Jose Maria Gonzalez: *Mobile IP*, <http://www.cs.berkeley.edu/~adj/cs294-1.s98/notes5.html>, 03/15/2003
- [10] RFC 791: *Internet Protocol*, <http://rfc.net/rfc791.html>, 03/15/2003
- [11] Mark Norris: *Mobile IP - Technologie for M-Buisness*, Artech House Inc., Norwood, 2001
- [12] IEEE 802.11: *Wireless Standard Package*, ISBN 0-7381-2315-3, 1999
- [13] European Telecommunications Standards Institute (ETSI): *High Performance Radio LAN2 (HiperLAN2)*, <http://www.hiperlan2.com>, 03/15/2003
- [14] U.Gärtner, F. Gonsior, D. Heerlein: *Design and Implementation of a Lab Course for Mobile Systems*, University of Federal Armed Forces Munich, Germany, Information Systems Laboratory (IIS), 2002

- [15] B. Stiller, G. Fankhauser, B. Plattner, N. Weiler: *Pre-study on Customer Care, Accounting, Charging, Billing, and Pricing*, Institut für Technische Informatik und Kommunikationsnetze TIK, ETH Zürich, February 18, 1998
- [16] WooMS e.V., <http://www.wooms.net>, 03/15/2003
- [17] WooMS e.V.: *The Organisation*, <https://www.wooms.net/phpnuke/html/modules.php?name=Content&pa=showpage&pid=5>, 03/15/2003
- [18] Leibniz-Rechenzentrum: *Kommunikation mit mobilen Rechnern im MWN*, <http://www.lrz-muenchen.de/services/netz/mobil/>, 03/15/2003
- [19] Leibniz-Rechenzentrum: *FunkLAN Accesspoints im Münchner Wissenschaftsnetz*, <http://www.lrz-muenchen.de/services/netz/mobil/accesspoints/>, 03/15/2003
- [20] ZD-Net: *Zukunft des Wireless LAN*, http://www.zdnet.de/mobile/artikel/techreport/mobile-business/mobile-business05_03-wc.html, 03/15/2003
- [21] Charles E. Perkins: *Mocile IP Design Principles and Practices*, Addison-Wesley, 1998, ISBN 0-201-63469-4
- [22] John Farserotu, Ramjee Prasad: *IP/ATM Mobile Satellite Networks*, Artech House Boston, 2002, ISBN 1-58053-112-1
- [23] Mark Norrris: *Mobile IP Technology for M-Buisness*, Artech House Boston, 2001, ISBN 1-58053-301-9
- [24] Jochen Schiller: *Mobilkommunikation*, Addison-Wesley, 2000, ISBN 3-8273-1578-6
- [25] Tero Ojanperä: *Wideband CDMA*, Artech House Boston, 2001, ISBN 1-58053-180-6

Kapitel 2

Analysis of and Outlook for the Mobile Commerce Sector

Uwe Gärtner

Noch vor 20 Jahren wusste kaum jemand etwas mit dem Begriff Mobiltelefon anzufangen. Zu Zeiten des analogen Festnetzanschlusses als hauptsächliches Telekommunikationsmittel war nicht vorzustellen, dass in nicht allzuferner Zukunft ein Großteil der Bevölkerung ständig und überall, nicht nur telefonisch sondern auch per SMS, MMS, eMail u.s.w., erreichbar ist.

Wenn man von mobile Commerce spricht meint man in erster Linie die Dienste und Services, die mit den neuen Technologien in der Mobilkommunikation an Relevanz gewinnen. Aus diesem Grund wird hier mit einem Überblick über die Entwicklung der Mobilkommunikation begonnen. Im Anschluß beschäftigt sich diese Seminararbeit mit den zukünftigen Anwendungsbereichen im M-Commerce und den in diesem Segment wichtigen Zahlungsverfahren. Zum Schluß wird versucht, mit den gegebenen Risiken und Erfolgsfaktoren, einen Ausblick in die Zukunft des M-Commerce-Sektors zu geben.

Inhaltsverzeichnis

2.1	Vom analogen Mobiltelefon zum heiß umkämpften Handy-	
	markt	33
2.1.1	Die Generationen von Mobilfunknetzen	33
2.1.2	Weitere mobile Technologien	35
2.1.3	Dienste/Services/Protokolle	36
2.2	M-Commerce heute	37
2.2.1	Definition von „M-Commerce“	38
2.2.2	Vorteile von M-Commerce gegenüber dem klassischen E-Commerce	38
2.2.3	Anwendungen	40
2.2.4	Zahlungsverfahren	42
2.3	Wie es weitergeht	44
2.3.1	Erfolgsfaktoren und Risiken	44
2.3.2	M-Commerce auf dem Weg in eine bessere Zukunft?	46
2.3.3	I-Mode — Eine Erfolgreiche Anwendung	47
2.4	Zusammenfassung und Ausblick	49

2.1 Vom analogen Mobiltelefon zum heiß umkämpften Handymarkt

In diesem Kapitel soll die Entwicklung im Bereich der Mobilkommunikation betrachtet werden. Dabei wird hier vor allem auf die verschiedenen Generationen von Mobilfunknetzen und die jeweils genutzten/ angebotenen Anwendungen und Dienste eingegangen.

2.1.1 Die Generationen von Mobilfunknetzen

Der Grundstein für die Mobilkommunikation wurde schon Anfang des letzten Jahrhunderts gelegt. Damals konnten erstmals Schiffe funktechnisch erreicht werden. In der weiteren Entwicklung wurde es auch möglich Fahrzeuge zu Luft und Land per Funk zu erreichen. Bereits 1950 gab es in Deutschland erste Inselnetze für Mobilfunk, diese arbeiteten im 30- bzw. 80-MHz-Bereich. Ende der sechziger, Anfang der siebziger Jahre wurden dann in Europa die ersten zellularen Mobilfunknetze installiert.[3]

- **A-Netz**

Das erste zellenbasierte Mobilfunknetz in Deutschland war das A-Netz. Es war ein Netz mit analoger Übertragung, das im Frequenzbereich von 156 MHz bis 174 MHz betrieben wurde. Wobei die Vermittlung noch per Hand vorgenommen wurde. Anfang der siebziger Jahre war es weltweit das größte zusammenhängende Mobilfunknetz mit zuletzt ca. 10000 Teilnehmern. 1977 wurde der Betrieb eingestellt.[3]

- **B-Netz**

Ebenso wie das A-Netz war auch das B-Netz ein analoges Mobilfunknetz, welches Anfang der siebziger Jahre installiert wurde. Jedoch wurden die Teilnehmer hier vollautomatisch vermittelt. Es arbeitete im Frequenzbereich von 146 MHz bis 156 MHz. Ein Handover war, wie es bei heutigen Zellularnetzen üblich ist, noch nicht möglich. Der Anrufer musste also wissen wo sich der Teilnehmer befand. Mitte der achtziger Jahre erreichte das Netz sein höchste Teilnehmerzahl von 25000 und wurde 1994 stillgelegt.[3]

- **C-Netz**

Auch das C-Netz arbeitete analog, allerdings erfolgte die Signalisierung beim Aufbau der Verbindung digital. Es war ein zellulares Netz im 450-MHz-Bereich, das national begrenzt 1981 in Betrieb genommen wurde. Im Gegensatz zum A- und B-Netz verfügte das C-Netz über einen automatischen Handover. Die Nutzung beschränkte sich im wesentlichen auf die Sprachübertragung, da die Datenübertragung den qualitativen Ansprüchen nicht entsprach. Erstmals wurde eine Teilnehmer-Berechtigungskarte eingesetzt, diese ermöglichte dem Teilnehmer die Nutzung des Netzes ohne an ein spezielles Endgerät gebunden zu sein. An Diensten bot das

C-Netz neben dem Handover die Rufumleitung und die Sprachspeicherung. Bedingt durch den Aufbau der modernen GSM-Netze konnte das C-Netz den gestiegenen Anforderungen nicht mehr gerecht werden und wurde Ende des Jahrhunderts abgeschaltet.[3]

A-, B- und C-Netz zählen auf Grund der analogen Sprachübertragung zu den Mobilfunknetzen der ersten Generation (1G). Mit Beginn der neunziger Jahre wurde erstmals die mobile Satellitenübertragung kommerziell genutzt. Außerdem wurden die analogen Übertragungstechniken durch digitale Signal- und Datenübertragung ersetzt, was die Anwendungen nicht mehr nur auf die Sprachübertragung beschränkte, sondern vielfältige Dienste und Services erlaubte und erlauben wird.

- **GSM**

Das „Global System for Mobile Communication“ (GSM) beschreibt zellulare Mobilfunknetze der zweiten Generation (2G). Mit GSM war es möglich die verschiedenen analogen Netze in Europa unter einem digitalen Standard zu vereinen. Inzwischen ist GSM der weltweit führende Mobilfunkstandard und ist mit hunderten Netzen in fast allen Ländern verfügbar. Der Standard arbeitet im 900-MHz-Bereich bzw. im 1,8-GHz-Bereich. Es stehen 9600 bit/s zur Datenübertragung zur Verfügung und im Gegensatz zu den analogen Netzen hat GSM eine hohe Sprachqualität und eine gute Fehlerkorrektur. Die Grunddienste sind in Teledienste und Trägerdienste aufgliedert. Wobei Teledienste die Nachrichtenübermittlung sicherstellt und garantiert das die Kompatibilität zu den Diensten der Festnetze gewährleistet ist. Trägerdienste sind Datentransportdienste die von den Betreibergesellschaften angeboten werden. Außerdem verwendet der Standard verschiedene Sicherheitsmechanismen (Authentisierung, Autorisierung, Verschlüsselung).[3][8]

- **GPRS**

Der „General Packet Radio Service“ (GRPS) dient der Datenübertragung in GSM-Netzen und soll eine Datenübertragung mit bis zu 171 kbit/s ermöglichen. Die Datenübertragung erfolgt hierbei in Datenpaketen und Kosten entstehen für den Nutzer nur wenn Pakete übertragen werden. GRPS gilt als Übergangslösung zum UMTS.

Weitere Standards, die ähnlich GRPS in GSM-Netzen eingesetzt werden, sind HSCSD (High Speed Circuit Switched Data) mit Übertragungsraten bis 76,8 kbit/s und EDGE (Enhanced Data Rates for GSM) mit bis zu 384 kbit/s.[3][7]

- **UMTS**

Das „Universal Mobile Telecommunication System“ (UMTS) ist ein von der ETSI 1998 standardisiertes System für die universelle Mobilfunktelekommunikation. Der Standard ist das Mobilfunknetz der dritten Generation (3G) und soll die bestehenden Netze der 2G mit einem erweiterten Leistungsspektrum ablösen. Gesendet wird in den Frequenzbereichen zwischen 1,900 GHz und 2,025 GHz sowie zwischen 2,110 GHz und 2,200 GHz. Mit einer Übertragungsrate von bis zu 2 Mbit/s (in einer sehr

kleinen Zelle) können dann Sprache, Daten, Faxe, Bilder, Videos und andere Daten übertragen werden. Die Vermittlungsstellen des UMTS-Netzes sollen durch ein ATM-Netz verbunden werden. Durch den ergänzenden Einsatz von satellitengestützten UMTS-Mobilfunksystemen soll die globale Erreichbarkeit für UMTS-Teilnehmer realisiert werden.[3][7][8]

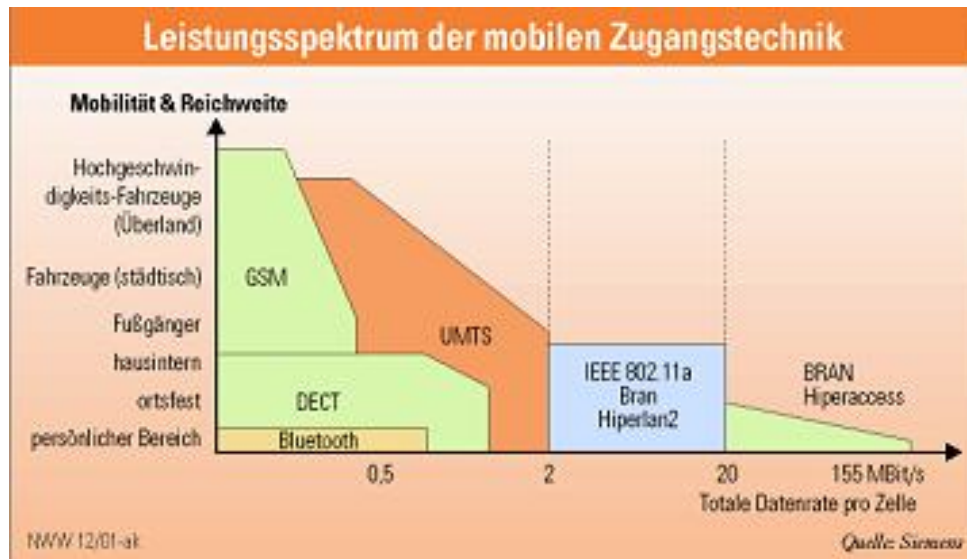


Abbildung 2.1: Mobilität und Übertragungsrate

2.1.2 Weitere mobile Technologien

In diesem Bereich fallen vor allem die wireless Networks, welches zumindest bei bestimmten Anwendungen eine Konkurrenz zu den Mobilfunknetzen bildet. Der größte Vorteil besteht wohl darin, daß sie in lizenzfreien Frequenzbereichen arbeiten und somit wesentlich kostengünstiger eingesetzt werden können.

- **Bluetooth**

Mit diesem Übertragungsstandard lassen sich vor allen Dingen lokale Daten austauschen und synchronisieren. Im Unterschied zur Infrarot-Technik ist dabei kein Sichtkontakt der Geräte notwendig. Die Entfernung kann 10 Meter betragen und es wird eine Geschwindigkeit von bis zu 1 mbit/s erreicht. Bluetooth eignet sich besonders für die Anbindung von Peripheriegeräten (Drucker, PADs, Laptops, Digitalkameras,...) an den PC.[1]

- **wLan**

Immerhin 11 mbit/s erlauben die Verbindungen per wLAN (wireless Local Area Network) heutzutage in den meisten Fällen. Am weitesten verbreitet sind derzeit Anwendungen auf der Basis der Spezifikation IEEE 802.11b, die im lizenzfreien 2,4

GHz Spektrum arbeiten. Damit ist eMail, File Transfer und Web Browsing ebenso wie die Interaktion mit „normalen“ LANs möglich. Der Standard der nächsten Generation ist 802.11a, allerdings nutzt er das 5 GHz Spektrum und ist damit in Europa nicht legal anwendbar. Ganz aktuell wird daher 802.11g die 25 mbit/s bieten.

Wenn es nach so namhaften Akteuren, wie z.B. Nokia und die Telekom mit ihrer US-Tochter VoiceStream, geht, soll eine Kombination aus 802.11 und dem bestehenden GPRS Netzwerk den Kunden das geben, was sie von Wireless Services erwarten. Die Idee ist in öffentlichen Gebäuden, wie Flughäfen, Kongresszentren und Cafes, so genannte Hot-Spots (wLan Zugangspunkte) aufzustellen. Mit entsprechenden Geräten kann man dann die weit aus bessere Verbindung nutzen. Ist das wLan gerade nicht erreichbar, wird die bereits bestehende GPRS-Verbindung genutzt. Aber auch hier sind noch einige Probleme zu lösen, so fehlt es noch an Endgeräten und im Bereich wLan sind die Sicherheit und das Abrechnungsverfahren noch nicht in dem Maße gewährleistet, wie es die Mobilfunknetze bieten.[1]

2.1.3 Dienste/Services/Protokolle

Wenn man über Dienste und Services im Bereich der Kommunikation spricht, so wird sich kaum jemand finden der dabei an das alltägliche Telefongespräch denkt oder die damit im Zusammenhang stehenden Anwendungen wie das Anklopfen und die Konverenzschaltung. So oder ähnlich verhält es sich im Speziellen auch bei der Mobilkommunikation. Telefonieren und ständig erreichbar sein ist das normalste auf der Welt. Einwenig anders sieht es da schon bei der SMS (short message service) aus. Mit ihr ist das Handy nicht mehr nur das mobile Telefon, sondern hat sich in den letzten Jahren zu einem „Multimediakommunikationsmittel“ entwickelt. Ein entscheidender Faktor ist dabei die Steigerung der Übertragungsrate, mit der die Provider, anknüpfend an den Erfolg der SMS, weitere Applicationen auf den Markt gebracht haben und bringen werden. Im Folgenden sollen auf alte und neue Dienste, Services und Protokolle, die im Bereich des mobile Commerce eine Rolle spielten oder spielen werden, näher eingegangen werden.

- **SMS**

Der Kurznachrichtendienst erlaubt es, an Mobilfunkteilnehmer Nachrichten mit 163 Zeichen zu senden. Auf dem SMS-Dienst basierend sind weitere Anwendungen möglich, so lassen sich auch Mitteilungen von einem E-Mail-System zu einem Handy austauschen. Die Übertragungsgeschwindigkeit liegt dabei bei 400 bis 500 bit/s. Kurznachrichten werden über die Handy-Tastatur eingetippt. Der Versand geschieht auf dem üblichen telefonischen Weg. Der große Vorteil: Eine kurze Nachricht per SMS zu verschicken, ist oft billiger als ein Gespräch. Deshalb erfreut sich dieser Dienst auch wachsender Beliebtheit: Etwa eine Milliarde Mitteilungen schicken die Deutschen pro Monat von Handy zu Handy.

- **MMS**

Multimedia Messaging Standard (MMS)- Ermöglicht den Versand von Texten, Melodien und Bildern. Dabei ist die Nachrichtenlänge, Gestaltung und Dateigröße einer

MMS nicht begrenzt. Auch Videosequenzen sind möglich. Diese werden mit dem neuen Mobilfunkstandard UMTS ihren Einzug feiern.

Dabei können Text mit gängigen Zeichensätzen und Schriften, Sprache AMR 16-kodiert (MP3, Midi und Wav geplant), Bilder als JPEG, GIF (89a oder 87a) oder WBMP (Jpeg 2000 geplant), Video als MPEG 4 (Simple Profile), Quicktime und ITU-T H.263, MMS-Streaming Bestandteile in der Nachricht sein.[13]

- **WAP**

Das Wireless Application Protocol(WAP) erlaubt den direkten Zugriff auf Internet-Inhalte vom Handy aus. Es benutzt als Codesprache WML (Wireless Markup Language) statt HTML. Grafiken werden über das Format WBMP (Wireless Bitmap) dargestellt. Der Benutzer eines WAP-fähigen Mobiltelefons kann von (fast) jedem Ort Informationen abrufen. Eher schleppend verlief bis heute die Verbreitung und Nutzung von WAP. Die Ursachen hierfür sind die unbefriedigende Übertragungsgeschwindigkeiten, zu kleine Displays, teure Dienste sowie eine umständliche Navigation. Dies trägt neben einer eher zögerlichen Verbreitung von wapfähigen Handys zu einer bislang eher mäßigen Penetrationsrate bei. Und selbst wer über ein entsprechendes Endgerät verfügt, wird damit noch nicht automatisch zum WAP-Nutzer. Umfragen zeigen, dass sich 51% der deutschen Internetnutzer, die ein WAP-Handy besitzen, nicht für Angebote auf WAP-Portalen interessieren und diese auch so gut wie überhaupt nicht nutzen würden.[1]

- **i-mode**

i-mode ist ein offener Standard, der auf der Programmiersprache iHTML, dem cHTML sehr ähnlich, basiert. i-mode basiert auf zwei wesentlichen Anwendungen, der i-mode Mail, die von Handy zu anderen Handys oder PCs gesendet und empfangen werden kann und zahlreichen Contentanwendungen, die von tagesaktuellen Nachrichten zu Politik, Wirtschaft und Sport über Dating/Chat und ortsbezogene Informationen, wie Stadtpläne und Routenplaner, bis zu Spiele, Funanwendungen und Erotik reicht. i-mode wird derzeit über GPRS angeboten und ermöglicht bereits heute einen Einblick in die mobile Welt von morgen. Dank GPRS ist das i-mode Handy „always online“, d. h. dauerhaft mit dem Internet verbunden. Dadurch sind neben sog. Pull-Diensten, bei denen der Nutzer aktiv Informationen anfordert, auch Pushdienste möglich, so etwa eine automatische Information des Nutzers über ein bestimmtes Ereignis.[1]

2.2 M-Commerce heute

Nachdem im ersten Kapitel vor allem auf Entwicklung, Technik und Dienste eingegangen wurde, beschäftigt sich dieses Kapitel näher mit dem eigentlichem Thema dieser Arbeit, dem M-Commerce. Es ist schon angeklungen, bei der Mobilkommunikation geht es immer mehr in Richtung multimedialer mobilen Welt und wenn es nach den Anbietern geht wird bald jeder seinen kleinen PC in der Tasche haben, denn für die meisten ist der heimische

PC in der Hauptsache nichts anderes als die Möglichkeit im Internet zu surfen, sich zu informieren oder mit Freunden zu kommunizieren.

2.2.1 Definition von „M-Commerce“

„Mobile Commerce ist ein Konzept zur Nutzung von Informations- und Kommunikationstechnologien zur mobilen Integration und Verzahnung unterschiedlicher Wertschöpfungsketten oder unternehmensübergreifender Geschäftsprozesse und zum Management von Geschäftsbeziehungen.“[6]

Eine etwas plastischere Definition könnte so lauten:

„Unter Mobile Commerce (kurz: M-Commerce) fallen die Bereiche des Electronic Commerce, die mit Unterstützung von Mobilfunkgeräten, PDAs (Personal Digital Assistants) und anderen mobilen Geräten orts- und zeitunabhängig über computerbasierte Netzwerke abgewickelt werden. M-Commerce ergänzt damit die Möglichkeiten der elektronischen Geschäftsabwicklung per Internet um eine mobile Komponente.“[5]

2.2.2 Vorteile von M-Commerce gegenüber dem klassischen E-Commerce

Derzeitiger E-Commerce auf Basis stationärer PC's besitzt einige Einschränkungen, die seine Verbreitung behindern. Stationäre PCs stellen für Endverbraucher in der Regel zunächst ein relativ hohes Investment dar und bedürfen zugleich einer gewissen Expertise in der Bedienung. Ein wichtiger Grund für Internetuser, die bislang E-Commerce-Dienstleistungen nicht genutzt zu haben, ist die Angst vor unsicheren Bezahlungsweisen und davor, dass die eigenen Daten für zweckfremde Aktivitäten mißbraucht werden. Zwar beschwerten sich die User über das hohe Maß an Informationsüberflutung und die fehlende Personalisierung des Online-Angebots, sie sind jedoch nur selten dazu bereit, Anbietern ihre persönliche Daten zu überlassen um personalisierte Angebote zu erhalten.

Im Vergleich zur Nutzung stationärer Endgeräte weisen mobile Endgeräte im Electronic-Business einige Vorteile auf. Die Kernvorteile werden im folgenden dargestellt werden. Dabei besitzt jede Eigenschaft das Potential, bestehende Problemfelder beim klassischen E-Commerce z.T. zu überwinden und darüber hinaus gänzlich neue Services entstehen zu lassen.[11]

Ortsunabhängigkeit:

Die „Überall-Verfügbarkeit“ mobiler Services ist der offensichtlichste Vorteil von mobilen Endgeräten. Mit diesen Geräten ist der Benutzer unabhängig von seinem Aufenthaltsort in der Lage, jederzeit Echtzeit-Informationen und Transaktionen durchführen zu können. Dadurch reduzieren sich z.B. Planungsprozesse des Konsumverhaltens; viele Transaktionen sind spontan dann möglich, wenn es dem Benutzer in den Sinn kommt. Vor dem Einkauf in der Innenstadt muss bspw. nicht mehr daran gedacht werden, da er online über mobile Endgeräte jederzeit verfügbar ist.[11]

Erreichbarkeit:

Mit einem mobilen Endgerät kann der Benutzer jederzeit und überall kontaktiert werden. Das eröffnet die Möglichkeit, zeitnah von Ereignissen informiert zu werden. Mit modernen Geräten ist es für den Nutzer jedoch ebenfalls möglich, seine Erreichbarkeit auf einen bestimmten Personenkreis zu begrenzen. Ein Aktienanleger kann sich durch zeitnahe Informationsdienste bspw. über kritische Kursbewegungen seiner Aktienpositionen informieren lassen und so rechtzeitig handeln.[11]

Lokalisierbarkeit:

Mit Technologien wie GPS, E-OTD oder COO, ist es möglich den genauen Standort des Benutzers zu bestimmen. Neben bereits bekannten Navigationsanwendungen werden durch die Lokalisierbarkeit vollkommen neuartige Servicequalitäten möglich sein. Mit einer Lokalisierungsfunktion wird der WAP-Stadtplanservice zum Navigationssystem. Auf diesen aufsetzend können Shoppingrouten vorgeschlagen oder das nahegelegenste italienische Restaurants gesucht werden.[11]

Sicherheit:

Handys besitzen heute bereits standardmäßig die Möglichkeit, sogenannte SIM und Smart-Cards zu integrieren. Dies sind Mikrorechner in Kartenform mit einer sehr begrenzten Schnittstelle nach außen. Mit ihnen ist eine Authentifizierung von Personen möglich und bietet dadurch vielfältige Einsatzmöglichkeiten für den E-Commerce auf Basis einer sehr viel höheren Sicherheit als heute im stationären Zugang üblich. So kann ein Handy zur Brieftasche, zum Mitgliedsausweis und zum Haustürschlüssel werden.[11]

Convenience:

Handy's sind einfach zu bedienen und haben eine weitreichende Verbreitung und Akzeptanz gefunden. Die damit einhergehende im Vergleich zu komplexen stationären Geräten geringere Hemmschwelle bahnt den Weg für neue Dienste auf mobilen Endgeräten. Ein mobiles Telefon zum Surfen findet weit mehr Akzeptanz als ein stationärer Internet-PC zum Telefonieren.[11]

Sofortige Verfügbarkeit:

Mobile Endgeräte sind so aufgebaut, dass sie nach dem Einschalten sofort benutzt werden können. Im Gegensatz zu stationären PCs müssen sie nicht minutenlang hochgefahren werden (no boot-time). Spätestens mit der Einführung von GPRS wird der Verbindungsaufbau mit dem Internet via WAP kaum noch bemerkbar sein. Das surfen im Internet wird sich von der Bequemlichkeit her nicht mehr vom Blättern im lokalen Telefonregister unterscheiden.[11]

Kostengünstigkeit:

Im Vergleich zum stationären PC werden Mobile Endgeräte genauer auf ihren Verwendungszweck und damit komplexitätsärmer zugeschnitten, wobei die gesamte Mikroelektronik so hoch integriert werden kann, dass Sie auf einen Chip zusammengefasst sehr preiswert hergestellt werden kann. Im Vergleich zum PC werden daher nur die Teile gekauft, die wirklich gebraucht werden. Die vergleichsweise niedrigen Kosten wird sich positiv auf die Akzeptanz der Produktvariante „Handy“ auswirken.[11]

Personalisierung:

Mobilfunkteilnehmer sind eindeutig aufgrund ihrer persönlichen Rufnummer identifizierbar. So wie ein ISDN-Telefon überträgt auch ein Handy die Telefonnummer an die Gegenstelle, beispielsweise einen WAP-Server. Dieser kann dann, wie in modernen ISDN-basierten Call-Centern, personalisierte Angebot bereithalten, wodurch ein wahres One-to-One Marketing ermöglicht wird. Zwar existieren auch für stationären Internet-PC's bereits einige Ansätze, um eine User-Identifizierung zu realisieren, jedoch behindert die fehlende Standardisierung die allgemeine Verbreitung. Ein Mobiltelefon ist hingegen aufgrund seiner Rufnummer per se personalisiert.[11]

2.2.3 Anwendungen

- **Formen des M-Commerce**

- M-Financial Services (Mobile-Banking, Mobile-Broking)
- M-Security Services (digitale Unterschrift, sicheres Bezahlen)
- M-Shopping (Mobile-Ticketing, Mobile-Auctions, Mobile-Retail)
- M-Advertising (zeitgruppenorientierte Werbung)
- M-Entertainment (Mobile-Gaming, Mobile-Music, Mobile-Betting)
- M-Telematics (Verkehrsinfos)

- **Consumer Services**

Die Entwicklung von Consumer Services ist im vollen Gange, doch es besteht eine Unsicherheit, welche Applikationen Erfolg haben werden. Hersteller sind jetzt unter Druck geraten die neuen Applikationen schnell auf dem Markt zu präsentieren, doch auf der anderen Seite müssen sie sich mit Einschränkungen bei der noch immer sehr limitierten Bandbreite und der schlechten Verfügbarkeit von neuen Endgeräten herumschlagen.

Die Consumer Services lassen sich in vier Segmente unterteilen: Informationen, Kommunikation, Unterhaltung und Transaktion.[4]

Informationen:

Die digitale Verbreitung von Informationen ist sehr stark am wachsen und kann auch über mobile Endgeräte verbreitet werden. Beispiele solcher Services sind Sportnachrichten, Wetterdienste, Aktienkurse, Verkehrsnachrichten und weitere.

Das Segment Informationen wird wahrscheinlich keine großen Umsätze erzielen, obwohl eine starke Personalisierung möglich ist, weil Konsumenten schon heute nicht bereit sind für solche Services im Internet zu bezahlen.[4]

Kommunikation:

Mobile Massaging Applikationen gelten als „Killerapplikationen“ im UMTS Umfeld und beinhalten person-to-person Kommunikation ohne Sprachdienste. Diese Vorhersage basiert auf dem enormen Erfolg von SMS in der mobilen Telekommunikation und von E-Mail und Instant Massaging in der Internet Welt.

Das Mobiltelefon kann als neues Medium zur Werbung eingesetzt werden und unterscheidet sich besonders in einem Punkt von herkömmlichen Werbemedien wie TV, Zeitschriften und dem Web. Durch die Möglichkeit einer Identifizierung der mobilen Kunden kann M-Advertising individuell auf spezifische Kundengruppen angepasst werden.[4]

Unterhaltung:

Unterhaltungsdienste sind mobile Spiele, Wetter, Icons, Klingeltöne, Fotos, Audio und Video. Heute wird in diesem Marktsegment noch wenig Umsatz erzielt und ausser Klingeltönen und Icons werden fast keine anderen Dienste angeboten. Mobile Unterhaltung kann in ein paar Jahren ein großes Umsatzwachstum erzielen, wenn größere Bandbreiten und bessere Displays zur Verfügung stehen. In diesem Segment liegt sicher ein riesiges Potential.[4]

Transaktion:

M-Tailing, oder auch M-Shopping genannt, ist ein zusätzlicher elektronischer Verkaufskanal, und Applikationen, welche die Vorteile mobiler Kommunikation (Lokalisierung, Personalisierung) nutzen, werden hier erfolgreich sein. „Killerapplikationen“ in diesem Segment sind m-auctions, m-promotion und andere zeitkritische Verkaufsplikationen.

Im mobilen Finanzbereich entsteht für Banken ein neuer Distributionskanal für traditionelle und online Bankdienstleistungen. Finanzinstitute investieren große Beträge um einem eventuellen Wachstum im mobilen Finanzmarkt stand zu halten.[4]

• Business Services

Unternehmen unterstützen ihre Prozesse mit Applikationen um konkurrenzfähig zu sein, d.h. Kosteneinsparungen oder Umsatzsteigerungen. Mobile Applikationen haben ein Potential zur Optimierung und Erweiterung von bestehenden Geschäftsprozessen, weil sie überall und jederzeit zur Verfügung stehen. Im Business Service gibt es drei Segmente für mobile Applikationen: Customer Relationship Management (CRM), Supply Chain Management (SCM), Workforce Applications (mobile Office).[4]

M-CRM:

CRM Lösungen fördern die Kundenbetreuung und unterstützen die Unternehmen im Verkauf, Service und Marketing. Es erlaubt die Erweiterung bestehender Dienste und Transaktionen mit mobilen Endgeräten. Diese erweiterten Lösungen eröffnen Möglichkeiten für neue Produkte und Dienstleistungen.

M-SCM:

Die rasante Entwicklung von E-Business Lösungen und elektronischen Marktplätzen hatte einen großen Einfluß auf das SCM. Mit der Erweiterung bestehender Lösungen mit mobilen Applikationen kann die Effizienz noch weiter gesteigert werden.[4]

M-Workforce:

Mobile Workforce Applikationen vereinfachen die Geschäftsprozesse für mobile Mit-

arbeiter. Die Planung, Zusammenarbeit und Kommunikation zwischen Mitarbeitern wird erleichtert. Die Zeit der überzogenen Erwartungen rund um das mobile Internet ist vorbei. Jetzt geht es darum, auf den Boden der Realität die wirklichen Nutzen- und Umsatzträger zu identifizieren. Mobile Firmenportale gehören sicherlich dazu.[4]

2.2.4 Zahlungsverfahren

Ein wichtiger Aspekt, der wiederholt im Zusammenhang mit M-Commerce genannt wird, ist die Fähigkeit das mobile Endgerät auch als Finanzinstrument zu nutzen. Grundsätzlich lassen sich im M-Commerce sämtliche vorstellbaren Lösungen zur Bezahlung denken, die auch bisher und im klassischen E-Commerce genutzt werden (Bsp. Kreditkartenzahlung und Rechnungen). Allerdings zeichnet sich das mobile Endgerät durch eine Reihe von Besonderheiten aus, die auch Einfluß auf die Sicherheits- und Bezahlungsfunktionen haben. Einige davon werden in Folge kurz vorgestellt. Bei aller neuen Technologie sollte aber nicht vergessen werden, daß der Kunde relativ genaue Anforderungen und Wünsche an die zu benutzenden Zahlungswege hat. Nicht zu unterbewerten ist vor allem die klassische Überweisung nach Erhalt der Rechnung, denn dies ist weithin mit großem Abstand die meistgewünschte Zahlungsart.[4]

- **Technologien**

SIM-Karte (Subscriber Identification Module)

Der Nutzer ist durch seine SIM-Karte und durch die hierbei notwendige Eingabe einer PIN vor Inbetriebnahme des mobilen Endgerätes ausreichend identifiziert. Der Netzbetreiber kann genau sehen, wer wann wie lange sein Endgerät nutzt. Damit gewinnt der Provider ein bedeutendes Werkzeug. In Verbindung mit weiteren Sicherheitsstandards, wie zum Beispiel der PKI-Technologie, könnte eine individuell vereinbarte Datenspeicherung auf der SIM-Karte eine Möglichkeit zur Zahlungsauslösung bedeuten.[4]

PKI(Public/Private Key Infrastructure)

PKI ist eine neue Verschlüsselungstechnologie, die für jede Art der Datenübertragung verfügbar ist. Bei diesem Verfahren benötigt man für die Ver- und Entschlüsselung ein Schlüsselpaar, den Private- und Public-Key. Der Absender verschlüsselt die zu sendende Nachricht mit dem Public-Key. Diese kann von dem Empfänger nur mit dem, zum Public-Key passenden, Private-Key Entschlüsselt werden. Das Verfahren ist auch als asymmetrische Verschlüsselung bekannt. Mit dieser Methode ist es unter anderem möglich Dokumente (z.B. E-Mail) auf unsicheren Kommunikationswegen (z.B. Internet, WLAN) zu übertragen.[15]

Micropayments

Micropayments bergen ein großes Potential im M-Commerce. Die bevorzugte Lösung ist die Abbuchung über die Telefonrechnung. Noch sträuben sich aber insbesondere die Netzbetreiber gegen diese Lösung, da sie unversehens in die Rolle eines



Abbildung 2.2: Ablauf beim PKI-Verfahren

Inkasso-Unternehmens geraten können und viele der haftungsrechtlichen und Gewinnverteilungsfragen noch nicht geklärt sind.[4]

Dual Slot Endgeräte

Ein mobiles Endgerät hat dann neben der SIM-Karte einen weiteren Slot für eine zweite Karte, die bei Bedarf eingesteckt wird. Der fällige Betrag wird dann direkt von der Karte abgebucht. Allerdings gibt es noch keine serienreife Endgeräte und die Geldkarte hat sich bisher noch nicht durchgesetzt. An eine Realisierung der gleichen Funktionalität für EC- oder Kreditkarte wurde bisher noch nicht gedacht.[4]

• Bestehende Lösungen

„Buy and Pay“

Bei dieser Art der Zahlungsabwicklung soll der Kaufvorgang folgende Handlungen beinhalten:

Der Kunde meldet sich bei „Buy and Pay“ durch Angabe seiner Adresse und seines Kontos an. Danach erhält der Kunde einen Benutzernamen und ein Passwort. Der Kunde erhält dann vom „Buy and Pay“-Server eine temporäre Session-ID. Diese soll der Anonymisierung der Kunden gegenüber dem Anbieterserver als auch der Identifikation bei „Buy and Pay“-diensten. Mit dieser Session-ID kann der Benutzer nun bei den verschiedenen Anbietern seinen Warenkorb füllen. Dieser wird anschließend mit der Session-ID durch den Anbieter zu „Buy and Pay“ übertragen. Der Kunde hat nun noch einmal die Möglichkeit seinen Warenkorb zu überprüfen und bestätigt anschließend den Kauf. Die Kaufbestätigung geht jetzt zurück zum Anbieter, der nun mit der Lieferung fortfahren kann.[9]

„Genion M-Payment“

Kernstück der von VIRBUS erstellten Lösung ist der AccountServer, der als intelligentes Relais zwischen Payment Gateway bzw. Bank, Händler und Kunden vermittelt. Der zweistufige Kaufvorgang beginnt mit der Authentisierung des Teilnehmers durch dessen Mobiltelefon. Dazu erhält er z.B. eine SMS mit einer für kurze Zeit gültigen Transaktionsnummer (TAN). Bei Nutzung eines WAP-fähigen Handys entfällt dieser Zwischenschritt, so dass zur Zahlungsautorisierung lediglich die Eingabe der Kundenkennung (PIN-Code) erforderlich ist. Im AccountServer laufen die Fäden der Transaktion zusammen: nach erfolgter Identitätsprüfung gilt es, den Zahlungsauftrag an den Payment Gateway weiterzuleiten. Dieser stößt den eigentlichen Zahlungsvorgang an. Danach werden Händler und Kunde über den Erfolg der

Zahlung informiert und die gewünschte Ware oder Dienstleistung kann transferiert werden.[10]

„paybyte“

Im Internet-Shop wird eine Telefonnummer angezeigt, welche angerufen werden muß. Die angegebene Transaktionsnummer wird am Telefon eingegeben und der Betrag erscheint auf der nächsten Telefonrechnung.[12]



Abbildung 2.3: Ablauf bei „paybyte“

2.3 Wie es weitergeht

Anknüpfend an Kapitel 2.2, wird in diesem Abschnitt auf Risiken und Erfolgsfaktoren für die Mobilfunkanbieter eingegangen. Klar ist aber schon, dass bisher große finanzielle Anstrengungen unternommen wurden und weitere nötig sind um den Kunden die neue Technik zugänglich zu machen und dessen Interesse für neue Anwendungen und Dienste zu wecken. Auch die bisher erstellten Marktprognosen lassen keinen eindeutigen Blick in die Zukunft zu, so das man gespannt sein darf, wie und wohin sich die Mobilfunkbranche entwickeln wird.

2.3.1 Erfolgsfaktoren und Risiken

Mobile Applikationen und somit auch der gesamte M-Commerce Markt werden nur erfolgreich sein, wenn die Entwickler die Vorteile der mobilen Kommunikation nutzen. Als kritische Erfolgsfaktoren gelten Personalisierung, Lokalisierung, Allgegenwart und Benutzerfreundlichkeit.

Personalisierung

Durch Personalisierung werden Dienste und Applikationen den individuellen Kundenpräferenzen angepasst. Sie basiert auf einem one-to-one Relationship Management und ist ein ideales Werkzeug für one-to-one Marketing. Dabei spielen Kundeninformation wie

Adresse, Telefonnummer, E-Mail, Endgerätetyp, Anrufverhalten u.s.w. eine wichtige Rolle. Zusätzlich sollte es möglich sein aus dem Kundenverhalten zu lernen und die Dienste und Services diesem dynamisch anzupassen.[4]

Lokalisierung

Schon heute stehen die verschiedensten Lokalisierungstechnologien zu Verfügung um Mobilteilnehmer auf wenige Meter genau zu orten. So wird es möglich dem Kunden genau auf seinen Standort angepasste Dienste und Services anzubieten. Diese Dienste sind der Schlüssel, die dem Endkunden einen spürbaren Mehrwert vermitteln.[4]

Allgegenwart

Die Fähigkeit Informationen zu erhalten und Transaktionen durchzuführen unabhängig vom Standort ist vor allem für zeitkritische Applikationen, wie z. B. Börsengeschäft und Wetten, sehr wichtig.[4]

Benutzerfreundlichkeit

Technologie ermöglicht dem Benutzer nur ein besseres Leben, wenn sie benutzerfreundlich ist. Bei der Entwicklung neuer Applikationen und Endgeräten ist dieser Erfolgsfaktor von entscheidender Bedeutung, denn Neues wird nur dann gebraucht wenn es die Handhabung alltäglicher Tätigkeiten vereinfacht und beschleunigt.[4]

Zum gegenwärtigen Zeitpunkt ist das Kundeninteresse am M-Commerce als eher gering einzuschätzen. Nach der anfänglichen Euphorie, mit Einführung der neuen Netze (GPRS, UMTS, ...) und Applikationen (WAP, MMS,...), wurden Betreiber und Kunden schnell auf den Boden der Realität zurückgeholt. Die verschiedensten Gründe spielen hierfür eine Rolle. Zu geringe Übertragungsraten, wenige Angebote, hohe Preise, Sorgen über die Sicherheit beim Nutzer, umständliche Handhabung der Endgeräte und fehlende Personalisierung sind einige der Faktoren warum der Boom im M-Commerce bisher ausblieb.[5]

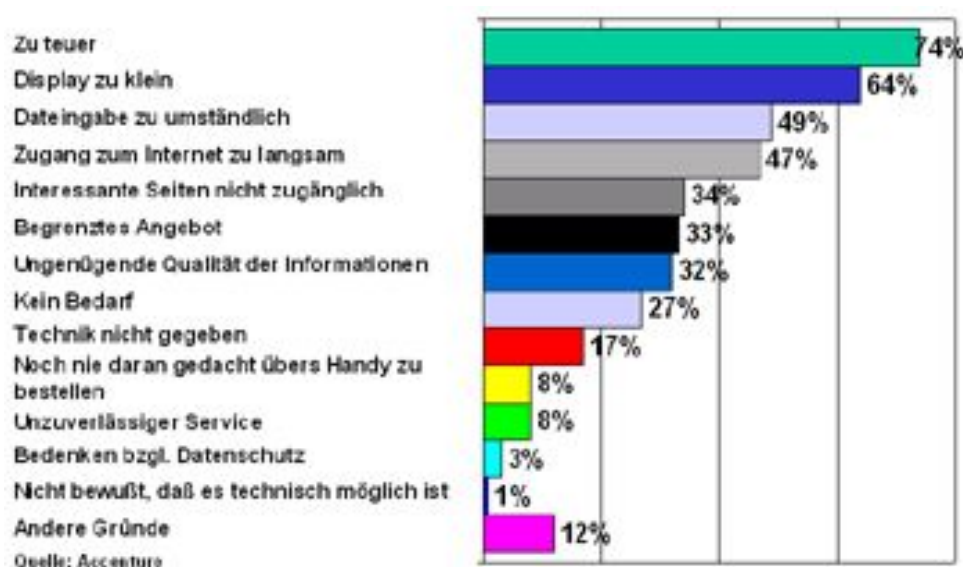


Abbildung 2.4: Hindernisse für M-Commerce

2.3.2 M-Commerce auf dem Weg in eine bessere Zukunft?

Die Begeisterung für mobile B2B (Business-to-Business) und B2C (Business-to-Consumer) Services hat weltweit nachgelassen. Aus diesem Grund gründete die Branche die „Open Mobile Alliance“ mit Mitgliedern aus führenden Unternehmen des Marktbereiches. Ziel dieser Vereinigung soll es sein, das Wachstum des Mobilien-Marktes voranzutreiben, indem Hemmnisse der globalen Nutzeradoption beseitigt werden, und die nahtlose Anwendungs-Kompatibilität sichergestellt wird.[2]

- **M-Commerce weltweit**

Die Prognosen im M-Commerce-Sektor gehen weit auseinander. So werden vom WAP-Forum Einnahmen in Höhe von 117 Milliarden USD für 2004 erwartet, dem gegenüber stehen 140 Milliarden USD die von Ovum prognostiziert werden. Vorallem die neuen mobilen Dienste sollen den Umsatz steigern. Damit verbunden wird allerdings auch sein, dass eine Verminderung bei den Sprachdiensten eintreten wird. Er soll von 93% in 2001 auf einen Anteil am Gesamtumsatz von 65% bis 2006 fallen. Die Gewinner sollen dann Massaging (14%), Infotainment (5%), Business Services (4%), sowie Werbung (4%) sein.

Am Beispiel des MMS soll nocheinmal klargestellt werden wie stark die verschiedenen Prognosen differieren. Für diesen Bereich erwartet Ovum 2007 ein Volumen von 70 Milliarden USD. Ganz anders schätzt das Wireless World Forum die Zukunft des MMS ein. Für 2006 erwartet das Forum weltweit ein Marktvolumen von 5,8 Milliarden USD. Diese gravierenden Unterschiede in den Vorhersagen zeigen die Unsicherheiten, die zur Zeit noch in diesem Marktsegment bestehen. Nur eines kann man den Zahlen entnehmen, der mobile Markt wird auf jeden Fall wachsen. Doch wohin genau die Entwicklung gehen wird bleibt unschlüssig.[2]

- **M-Commerce in Europa**

Auch bei den Prognosen für Europa gibt es teils eklatante Unterschiede.

Ovum geht davon aus das 2005 42% (174,6 Millionen) der Bevölkerung M-Commerce nutzen. Damit einher geht auch der Anstieg der Umsätze von 19,59 Milliarden EUR auf 86,19 Milliarden EUR in 2005. Anders als im E-Commerce soll der Anteil der Privatkunden den weitaus größeren Kundensektor im M-Commerce ausmachen.

Die Erwartungen von Ovum werden von Analysys Resaerch noch übertroffen, denn hier geht man von 134,5 Milliarden EUR für 2007 aus. EITO ist noch optimistischer und erwartet 160,8 Milliarden EUR in 2006. Dagegen ist die Yankee Group weitaus weniger zuversichtlich. Es wird mit nur 53 Milliarden EUR Einnahmen für das Jahr 2006 gerechnet.

Auch Frost & Sullivan geben positive Erwartungen, hier in Bezug auf M-Payment, ab. Das Transaktionsvolumen soll von 0,3 Milliarden EUR (2002) auf 27,3 Milliarden EUR (2006) anwachsen. Dabei ist einer der Hauptgründe für die positive Beurteilung, dass die Sicherheit bei der Zahlungsabwicklung mit M-Payment höher ist als mit der Kreditkarte.

Die Umsatzerwartungen für den Mobile Content sind ebenfalls sehr positiv. Die

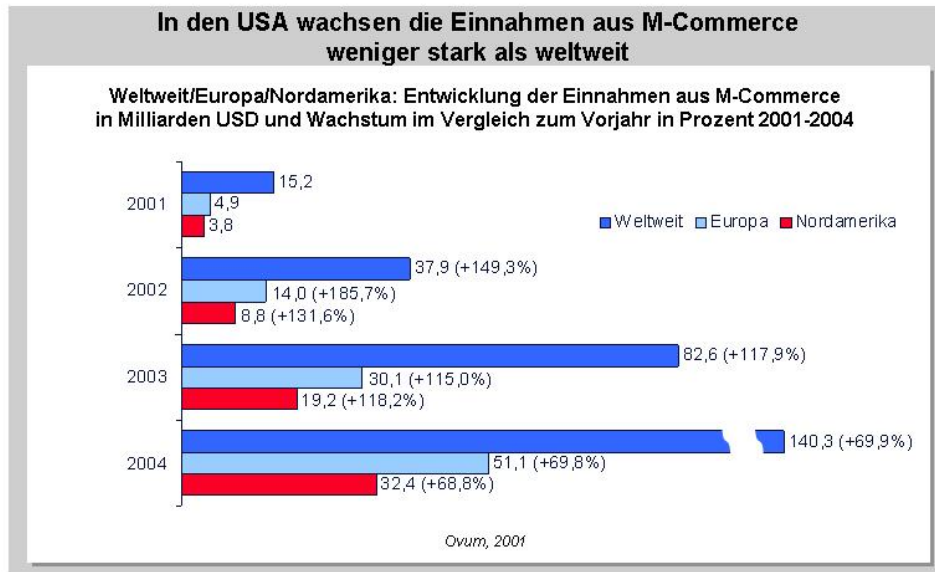


Abbildung 2.5: Entwicklung der Einnahmen aus M-Commerce

Analysten von Andersen prognostizieren für 2006 18,9 Milliarden EUR, eine Steigerung zu 2002 von 845%. Der Hauptteil des Umsatzes wird dabei mit Spielen und Nachrichten erzielt werden. Dagegen erwartet Jupiter MMXI nur 3,3 Milliarden EUR. Jedoch ist damit die Zahlungsbereitschaft für M-Content immernoch doppelt so hoch wie für E-Content.

Seit 2002 werden neue Dienste über MMS angeboten. Die Prognosen sehen auch hier mehr als positiv aus. Ovum prognostiziert einen Umsatz von 29 Milliarden USD für 2007. Die Umsatzsteigerung wird dabei zu Lasten des SMS gehen. Die Einnahmen werden von 18 Milliarden EUR (2002) auf 11,8 Milliarden EUR (2007) zurückgehen, wobei SMS der Anzahl nach immer noch dominierend bleibt.[2]

2.3.3 I-Mode — Eine Erfolgreiche Anwendung

Seinen Ursprung hat i-mode in Japan. Und im Land der aufgehenden Sonne hat sich dieser mobile Service inzwischen zu einem ebenso viel wie gern zitierten Paradebeispiel für erfolgreichen mCommerce entwickelt. Im Februar 1999 als exklusives Angebot der NTT DoCoMo gestartet, verfügt i-mode in Japan heute über 32,8 Millionen Nutzer, 2.004 lizenzierte Content Anbieter mit 3.005 offiziellen Sites sowie weiteren 53.725 so genannten „voluntary“ i-mode-Sites, die von unabhängigen Betreibern zumeist kostenlos angeboten werden. Laut Forrester bewegt sich mCommerce in Japan inzwischen in einer Größenordnung von 400 Millionen USD und übersteigt die aktuellen europäischen Umsätze damit um mehr als das 30fache.

Das eigentlich Besondere an i-mode liegt jedoch im Gesamtpaket begründet: Denn es handelt sich weitaus weniger um eine bloße technische Lösung als vielmehr um ein komplet-

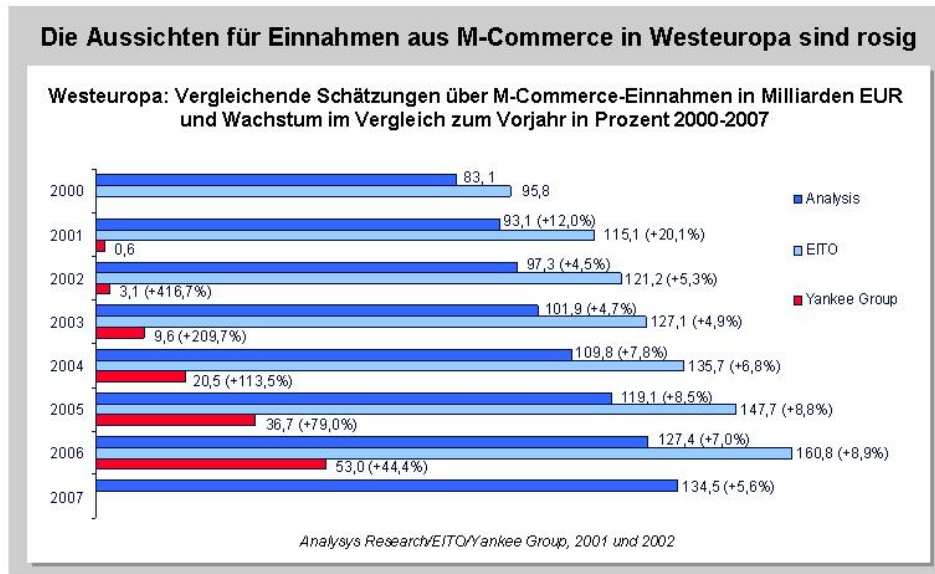


Abbildung 2.6: Vergleich der Schätzungen über M-Commerce-Einnahmen

tes Geschäftsmodell. Bei i-mode stand und steht vor allem die Bereitstellung eines breit gefächerten Content-Angebots im Vordergrund. Lizenzierte Anbieter stellen die Inhalte (Nachrichten & Infodienste, eMail, Ticketservices und Shopangebote, Spiele...) bereit, die nach Volumen abgerechnet werden und an denen der Serviceanbieter (also etwa NTT DoCoMo) wiederum mit einer Provision beteiligt ist. Für den Nutzer wird darüber hinaus zudem eine monatliche Grundgebühr fällig. Da sich Kosten und angebotene Leistungen in Japan in einem offenbar sehr günstigen Preis-Leistungsverhältnis bewegen, hat sich i-mode dort zu einem echten „Renner“ gemausert.

Ein wenig von diesem Erfolg partizipieren möchten in Europa die niederländische KPN sowie deren deutsche Mobilfunktochter ePlus. Gemeinsam bieten sie seit dem 16. März 2002 i-mode exklusiv in den Niederlanden und Deutschland an. Die Bilanz nach zwei Monaten: Bislang konnten 34.000 Kunden für den neuen Service gewonnen werden und die Zahl der Content-Anbieter hat sich von 60 Partnern beim Start auf inzwischen 90 erhöht. Daneben bestehen bis dato rund 700 private i-mode-Homepages, die mit den japanischen „voluntary“-Sites vergleichbar sind.

Von einem i-mode-Erfolg wie in Japan geht in Europa derzeit niemand aus. Neben dem Markt- und Entwicklungsvorsprung dürften vor allem auch das hiesige Kostenmodell sowie ein differierendes Verbraucherverhalten eher bremsend wirken. Hinzu kommen mögliche Versorgungsengpässe mit i-mode-fähigen Handys. Nachdem sich Nokia bislang beharrlich weigert, i-mode-Handys herzustellen, kann ePlus hierzulande nur mit Geräten von Nischenanbietern wie Nec oder Toshiba aufwarten. Positiv zu werten ist allerdings der mögliche „Lerngewinn“ beim Aufbau und der Vermarktung mobilen Contents. Denn wie bereits Yusuke Kanda von NTT DoCoMo im Rahmen der Euroform Conference in Düsseldorf feststellte: „Europäische Mobilfunk Anbieter sollten sich weniger Gedanken über die

Technologie als vielmehr über die anzubietenden Services machen, wenn sie beabsichtigen mit neuen breitbandigen Internet-Services Geld zu verdienen.“[1]

2.4 Zusammenfassung und Ausblick

Mit den ersten analogen Netzen vor 30 Jahren begann die Erfolgsstorie der Mobilkommunikation. Anteil an Mobilfunkanschlüssen in der Bevölkerung ist mit 58,5% (2000) genau so hoch wie bei den Festnetzanschlüssen (59,3%)[7]. Dagegen hatte das damalige B-Netz mitte der achziger Jahre gerade 25000 Teilnehmer. Zudem sind klare Unterschiede in der Anwendung der mobilen und festen Anschlüsse zu erkennen. Während beim Festnetz vorrangig der Sprachdienst genutzt wird, spielen bei der Mobilkommunikation, vorallem bei der jungen Nutzergeneration, die Datendienste (SMS, WAP...) eine große Rolle. Die Analysten gehen davon aus, dass dieser Trend sich vortsetzen wird und prognostizieren dem Datendiensten einen Anstieg von 7% auf 37% bis zum Jahr 2006 [2].

Diese Zahlen zeigen das ungeheure Wachstumspotential im M-Commerce-Sektor. Die Prognosen mögen den ein oder anderen, angesichts der eher stagnierenden Entwicklung beim M-Commerce, verwundern, doch bietet der Marktbereich gerade gegenüber dem E-Commerce große Vorteile. Ortsunabhängigkeit, Erreichbarkeit, Lokalisierung und Sicherheit sind Faktoren die den mobilen User durchaus auch schon heute überzeugen. Leider wurden die Vorteile, welche die neuen Technologien mit GPRS, UMTS, MMS, WAP u.s.w. bieten, von den Anbietern noch nicht umgesetzt, denn geringe Übertragungsraten, umständliche Handhabung der Endgeräte, zu kleine Displays, zu teure und wenige Dienste sind nur ein paar Gründe warum es dem Nutzer schwerfällt sich in der Welt des M-Commerce wohlfühlen.

Nur wenn die Provider die angesprochen Mängel beheben, stärker auf die Bedürfnisse und Erwartungen des Nutzers eingehen und ihm das Gefühl geben, dass er durch die Nutzung des M-Commerce einen Mehrwert hat, werden die hochgesteckten Prognosen der Analysten eintreffen.

Meiner Meinung nach wurden die Anbieter, durch die eher negative Entwicklung beim M-Commerce, wachgerüttelt. Sie werden durch neue Marketingkonzepte und den Ausbau der Mobilfunknetze das Kundeninteresse erneut wecken können. Vorallem da es ja mit i-mode ein erfolgreiches Beispiel gibt, das zeigt wie man es machen kann. So das die hochgesteckten Erwartungen für den M-Commerce-Sektor durchaus machbar sind, denn einmal ins rollen gekommen, wird sich die Welle, ähnlich wie es die SMS gezeigt hat, ausbreiten. In meinen Augen wird dabei der Hauptnutzeranteil im Busines-Bereich und den jüngern Usern liegen. Die größte Hürde die bei dieser Entwicklung zu nehmen ist sehe ich persönlich nicht in dem Ausbau der Netze oder Einführung von neuen Anwendungen, sondern das richtige Maß zwischen Kosten und Leistung zu finden. Dabei kommt es auch darauf an das die Mobilfunkbetreiber einsehen, dass sie nicht alleine auf dem Markt sind und mit verschiedenen Service-Providern zusammenarbeiten müssen, um die Atraktivität des Netzes und der Dienste für den Nutzer sicherzustellen.

Literaturverzeichnis

- [1] ecin, Trendbericht: Technik in der Telekommunikation,
<http://www.ecin.de/technik/standards/>
- [2] NFO Infratest, 5. Faktenbericht, <http://193.202.26.196/bmwi/index.htm> (11.2002)
- [3] NetworkWorld, Lexikon, <http://www.networkworld.de/>
- [4] Prof. H. Schauer, Dr. R. Riedl, M-Commerce, Universität Zürich (2001)
- [5] ECC, Mobile Commerce - Problemfeld,
http://www.ecc-handel.de/branchen_prob/968752926/
- [6] D. Pfaff, M-Commerce, www.infomarketing.de (2002)
- [7] Stars-Bytes, Mobile Kommunikation,
http://www.stars-bytes.de/3_lernen/3_ges/cont/8_com/2_mob.htm#1
- [8] N. Kaiser, Mobilkommunikation als Integrationsplattform(2000)
- [9] buy and pay, <http://www.buy-and-pay.de/funktion.html>
- [10] Virbus, Genion M-Payment, <http://www.virbus.de/de/projects/mpayment/>
- [11] F. Buckler, H. Buxel, Mobile Commerce Report,
http://www.profit-station.de/star/study2/sty02a_g.htm (2000)
- [12] paybyte, <http://www.paybyte.de/v2/>
- [13] 4phones, Lexikon, <http://www.4phones.de/lexikon/m.shtml>
- [14] E. Pech, M. Chiesa, Online Business 2002 and beyond, Einführungsvortrag,
Diebold Management und Technologieberatung GmbH (2002)
- [15] abylonsoft, Lexikon,
<http://www.abylonsoft.de/lexikon.htm#Asymmetrische%20Verschl%C3%BCsslung>

Kapitel 3

Peer-to-Peer: Die Betrachtung der Wirtschaftlichkeit verteilter Systeme

Florian Gonsior

Das Internet hat sich in den letzten Jahrzehnten explosionsartig entwickelt und es gibt kaum jemanden, der sich vor diesem Medium verschließt. Die ständig steigende Anzahl der Nutzer führte zu einer sehr hohen Auslastung des Internets. So kommt es z.B. bei vielen Servern in den Stoßzeiten zu langen Wartezeiten, welche für den Nutzer unbefriedigend sind.

Wie man sich leicht vorstellen kann, stößt das Client/Server-Modell unter diesen Bedingungen an seine Grenzen. In dieser Seminararbeit wird Peer-to-Peer (P2P) als mögliche Alternative zu einer auf Client/Server (C/S) basierenden Architektur betrachtet. Es bietet mit seiner Technik eine besondere Robustheit durch Dezentralisation. Neben der Entstehung von P2P-Netzen wird ein Vergleich zwischen den beiden Architekturen vollzogen, was sowohl die Struktur und die Technologie als auch die wirtschaftlichen Hintergründe und Aspekte betrifft. Desweiteren werden die Anwendungsbereiche näher betrachtet und die Abläufe der P2P-Kommunikation.

Inhaltsverzeichnis

3.1	Einleitung	53
3.1.1	Entstehung von P2P	53
3.1.2	Hauptkategorien der P2P-Anwendungen	53
3.1.3	Derzeitige Marktsituation für P2P-Anwendungen	55
3.2	Vergleich zwischen P2P und der C/S-Architektur	56
3.2.1	P2P im Vergleich zu C/S	56
3.2.2	Beispiel für den Nutzen von P2P	60
3.2.3	Kosten, Nutzen und Zukunft	60
3.3	Komponenten in einem P2P-System	62
3.3.1	Peer-Typen	62
3.3.2	Peer Groups	64
3.3.3	Network Transport	64
3.3.4	Services	65
3.3.5	Protokolle	66
3.4	Kommunikation im P2P-Netzwerk	66
3.4.1	Suche nach Peers und Services	67
3.4.2	Suche nach Rendezvous und Router Peers	68
3.4.3	Überwinden der Grenzen privater Netzwerke	68
3.4.4	Vorteile der P2P-Kommunikation	70
3.5	Zusammenfassung und Ausblick	71

3.1 Einleitung

Dieses Kapitel befaßt sich zunächst mit der Entstehung von P2P-Systemen, stellt anschließend die drei wichtigsten Anwendungsbereiche dieser Architektur dar, in welche sich fast alle P2P-Anwendungen eingliedern lassen und liefert einen Einblick in die derzeitigen Marktchancen für P2P-Anwendungen.

3.1.1 Entstehung von P2P

Die Frage nach dem Zeitpunkt des ersten Erscheinens von P2P muß man in die Frage nach der Entstehung eines reinen P2P-Netzwerks und die Frage nach dem Auftreten von P2P-Funktionalitäten unterteilen. Man liegt z.B. mit der Annahme falsch, daß P2P eine Neuentwicklung ist. P2P-Funktionalität hat es, eingebettet in das C/S-Prinzip, schon immer im Internet gegeben. So haben z.B. alle Server mit fester oder dynamischer IP schon immer die Möglichkeit gehabt, mit anderen Servern zu kommunizieren, bzw. deren Dienste in Anspruch zu nehmen.

Als die wohl erste wirkliche P2P-Anwendung, welche ein reines P2P-Netzwerk bildete, läßt sich Usenet [25] nennen. Es wurde 1979 von den North Carolina Studenten Tom Truscott und Jim Ellis entwickelt. Usenet sollte Computern einen Weg bieten, um Informationen auszutauschen. Dies geschah zu einer Zeit, wo es noch keine großflächige Internetverbindung gab, die man dazu hätte nutzen können. Aus diesem Grund gaben die beiden Studenten dem Computer die Möglichkeit einen anderen Computer anzurufen, nach neuen Daten zu sehen und diese zu kopieren. Dies geschah nachts, um hohe Telefonrechnungen zu vermeiden. Dieses System entwickelte sich zu dem umfangreichen Newsgroupsystem was wir heute vorfinden. An Usenet lassen sich die wichtigsten Merkmale von P2P erkennen [01]:

- Usenet besitzt keine zentrale Steuerungseinheit
- Die Verteilung des Inhalts wird von jedem Knoten gesteuert
- Der Inhalt des Usenet-Netzwerks existiert teilweise oder im Ganzen als Replikat in den einzelnen Knoten

Usenet läßt sich nicht als Software oder als Netzwerk von Servern definieren, es ist einfach eine Methode wie Computer miteinander kommunizieren können. Diese Kommunikation geschieht mit Hilfe des Network News Transport Protokolls.

Nach diesem wohl ersten Auftreten von P2P folgten viele Anwendungen, welche nachfolgend behandelt werden, mit ähnlichen Merkmalen.

3.1.2 Hauptkategorien der P2P-Anwendungen

Wie bereits erwähnt, lassen sich die meisten Anwendungen die auf P2P basieren in drei große Kategorien gliedern. Diese drei Bereiche werden anschließend näher betrachtet und anhand von bekannten Beispielen erläutert.

• Instant Messaging

Als wohl bekanntestes Beispiel läßt sich an dieser Stelle ICQ [11], was für "I Seek You" steht nennen, welches seit 1996 auf dem Markt ist. ICQ gibt seinen Nutzern eine schnellere Möglichkeit der Kommunikation im Gegensatz zur E-Mail. Sobald das Gegenüber online gegangen ist, besitzt man die Möglichkeit eine direkte Nachricht zu senden, desweiteren hat man die Möglichkeit eines direkten Datentransfers. Man muß jedoch erwähnen, daß ICQ keine reine P2P-Anwendung ist, sondern eine hybride Lösung. Wie in der nachfolgenden Abbildung dargestellt, benötigt ICQ einen Server, um anzuzeigen, welche Benutzer zur Zeit online sind und um Interessenten zu informieren, wenn neue Nutzer beitreten. Alle anderen Kommunikationen unter den Nutzern laufen jedoch im P2P-Modus ab [10].

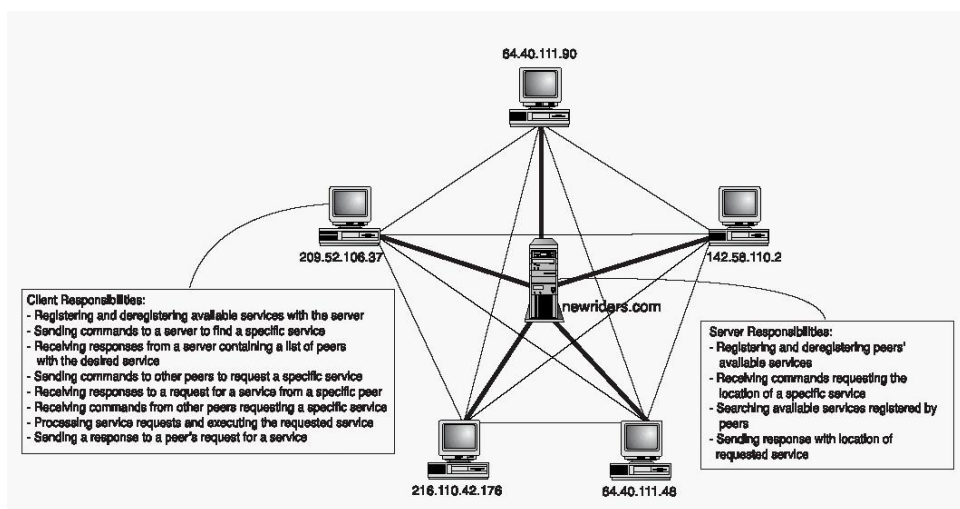


Abbildung 3.1: Hybride Peer-to-Peer Architektur [01]

Inzwischen hat ICQ viele Nachahmer. Zu den bekanntesten gehören der Yahoo! Messenger [13], der MSN Messenger [14] und der AOL Internet Messenger [15]. Aufgrund unterschiedlicher Protokolle sind die verschiedenen Messenger inkompatibel. Dies führt dazu, daß man sich jeden Messenger installieren müßte um Zugang zu allen Netzen zu haben. Es wuchsen mit der Zeit viele voneinander unabhängigen Nutzergemeinden.

Mittlerweile gibt es Software, welche sämtliche Protokolle unterstützt und somit die entstandenen Grenzen überwinden kann. Ein Beispiel für solche Programme ist Jabber [16]. Diese Programme stoßen allerdings auf Widerstand bei den Service Providern der Instant Messenger, so hat z.B. AOL sein Protokoll geändert [01].

• Distributed Computing

Distributed Computing ist ein sehr beeindruckendes Beispiel für die Möglichkeiten, die sich mit P2P erschließen. Es bietet einen Weg schwierige Probleme zu lösen, indem man das Hauptproblem in viele kleine Probleme zerteilt, welche dann unabhängig von einer großen Zahl von Computern bewältigt werden können. Bisher nutzen die Anwendungen zwar nur einen Teil der Möglichkeiten die P2P bietet, es ist jedoch anzunehmen, daß in naher Zukunft das volle Spektrum ausgeschöpft wird.

Zu den bekanntesten Projekten gehören wohl das SETI@Home Projekt [12] und Distribution.net [17]. Bei SETI@Home verarbeiten die Nutzer, mit Hilfe einer auf einem Bildschirmschoner basierenden Anwendung, Daten von einem Radio-Teleskop. Durch diese Applikation tragen sie einen Teil zur Suche nach außerirdischem Leben bei. Mittlerweile hat dieses Projekt mehr als drei Millionen Teilnehmer, von denen eine halbe Million aktiv sind.

Bei Distributed.net geht es darum, unlösbare Kodierungen zu entschlüsseln. Ähnlich dem SETI@Home Projekt laden sich die Teilnehmer einen Teil des zu lösenden Problems von einem Server herunter und agieren anschließend bis zur Lösung des Problems unabhängig.

- **File Sharing**

Die für den "normalen" Benutzer wohl reizvollste Anwendung von P2P ist das File Sharing. Ein bekanntes Beispiel ist Napster [18], welches ebenso wie ICQ eine hybride Architektur beinhaltet. Es existiert ein zentraler Server auf dem Listen der MP3 Dateien, welche auf den Rechnern der Nutzer existieren, abgelegt sind. Desweiteren verwaltet dieser Server die Rechte, ob man diese Listen nach Liedern durchsuchen und den zugehörigen Host sehen darf. Das eigentliche Kopieren der MP3 Dateien erfolgt dann mit einer P2P-Verbindung ohne Beteiligung des Servers.

Aufgrund der Rechtsverletzung in Bezug auf das Copyright verfeinerte Gnutella [19] den Ansatz von Napster hin zu einer reinen P2P Lösung. Es gibt also keine Notwendigkeit mehr für einen Server. Bei Gnutella bieten die einzelnen Rechner nicht nur ihre Dateien an, sondern sind auch verantwortlich für die Weiterleitung von Queries und Routing-Nachrichten. Auch wenn Gnutella keinen Server mehr braucht, um die Suchfunktion anzubieten, so ist es doch notwendig eine IP von einem Peer zu kennen, der bereits in das Netz integriert ist. Aus diesem Grund wurden einzelne Peers mit statischer IP errichtet, welche neuen Nutzern einen Startpunkt bieten sollen, um Zugang zum Netz zu erhalten.

3.1.3 Derzeitige Marktsituation für P2P-Anwendungen

Anhand der aufgeführten Beispiele ist sicherlich jedem aufgefallen, daß auch er schon einmal mit P2P in Kontakt gekommen ist. Als wohl interessantestes Beispiel für P2P muß man wohl Instant Messaging hervorheben. Ein Indiz für den Zuspruch auf der Nutzerseite sind die rapide wachsenden Mitgliederzahlen von ICQ. In einer Zeit, in der nahezu jeder einen Internetzugang besitzt, stellt Instant Messaging eine attraktive Form der Kommunikation dar. Bei ICQ wird z.B. die preiswerte Seite von E-Mails mit der schnellen Handhabung von SMS kombiniert.

Während Instant Messaging für alle Nutzer gleichermaßen reizvoll ist, so genießen die Vorteile von Distributed Computing doch eher die Firmen. Die Rechenleistung, welche mit dieser Technologie erreicht wird, ist sicherlich beachtlich und besonders kostengünstig, doch wird es nur vereinzelte Einsatzgebiete für Distributed Computing geben. Früher existierten weitere kleine Einsatzgebiete, welche so kostengünstig die benötigte Rechenleistung erreichten (z.B. digitale Animationen und Grafikberechnungen). Im Laufe der Leistungsexplosion auf der Hardwareseite wurde ein solches Verfahren für viele unnötig. File Sharing bildet eine Brücke zwischen Instant Messaging und Distributed Computing,

es ist sowohl für einzelne Nutzer attraktiv, als auch für große Unternehmen reizvoll. Während die Nutzer von der Verfügbarkeit aller erdenklichen Daten profitieren, kommt es den Unternehmen eher auf die kostengünstige und stabile Realisierung von Firmendatenbanken an. Auf der Nutzerseite gibt es jedoch eine große Gefahr in Sachen Rechtsverletzung. Ein P2P-Netzwerk bietet die schwer zu kontrollierende Möglichkeit, urheberrechtlich geschützte Daten zu verbreiten. Diese Rechtsverstöße wurden bereits zu einem gravierenden Problem für Napster, welches Listen über vorhandene Musiktitel weiterhin auf einem Server bereit stellte. Gnutella umgeht dieses Problem, indem kein Server mehr genutzt und somit die Kontrolle erschwert wird.

In Zukunft wird das größte Wachstum in den Bereichen Instant Messaging und File Sharing zu erwarten sein, während Distributed Computing auf vereinzelte eindrucksvolle Projekte beschränkt bleibt. Es bleibt abzuwarten, wie sehr der Zuwachs bei Instant Messengern gebremst wird, sobald diese nur noch kommerziell angeboten werden. Ebenso könnte der Gesetzgeber dem File Sharing klare Grenzen vorgeben (Probleme mit Copyright), welche die Attraktivität mindern.

Zur Zeit besteht allerdings eine große Nachfrage für P2P-Anwendungen, deren Ende nicht abzusehen ist.

3.2 Vergleich zwischen P2P und der C/S-Architektur

In diesem Kapitel werden die beiden Architekturen gegenübergestellt und miteinander verglichen. Dieser Vergleich erfolgt im ersten Teil allgemein und wird im zweiten Teil durch ein Beispiel veranschaulicht. Den Abschluß des Kapitels bildet ein tieferer Einblick in Zahlen und Fakten bezüglich Kosten, Nutzen und Leistungsfähigkeit.

3.2.1 P2P im Vergleich zu C/S

Die meisten Internet Dienste werden heute mit Hilfe der C/S-Architektur angeboten, welche in der nachfolgenden Abbildung veranschaulicht wird. Bei dieser Technologie bauen die Clients eine Verbindung zum Server mit Hilfe eines bestimmten Kommunikationsprotokolls (z.B. FTP) auf, um Zugang zu einer bestimmten Ressource zu erhalten. Der Großteil der Arbeit, bei einem solchen Zugriff auf irgendwelche Dienste, läuft auf der Serverseite ab. Die Clients bleiben dabei relativ ungenutzt.

Wichtige Beispiele für dieses Verfahren sind:

- World Wide Web
- FTP
- telnet
- E-Mail

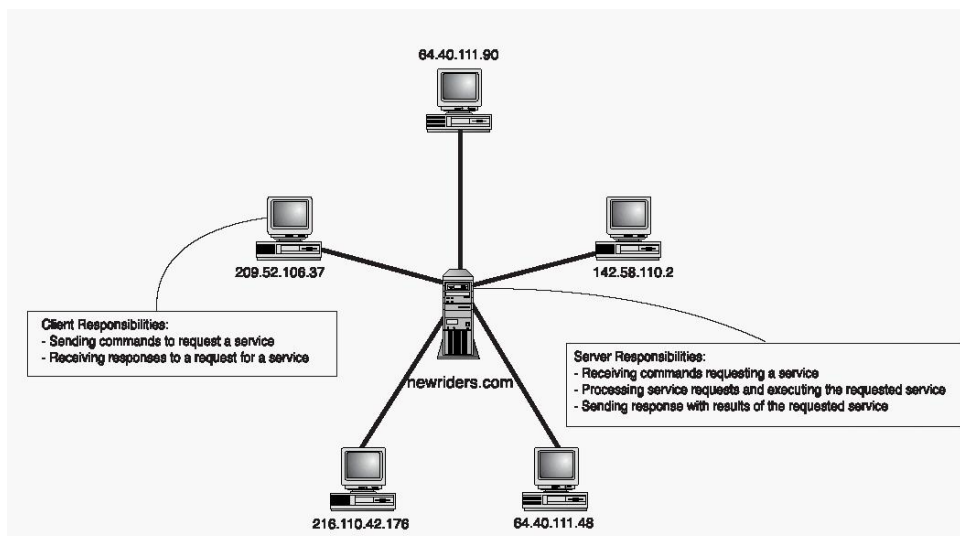


Abbildung 3.2: Client/Server Architektur [01]

Die C/S-Architektur hat jedoch einen entscheidenden Nachteil, so führt das Wachstum der Anzahl der Clients dazu, daß auch die benötigte Bandbreite und Last auf der Serverseite wächst. Dies könnte zur Folge haben, daß der Server keine weiteren Clients bedienen kann. Der Vorteil den dieses System bietet, ist in der heutigen Zeit immer bedeutungsloser geworden. Der Server nimmt den Client-Systemen durch die geringe Abgabe von Arbeit den Druck besonders leistungsfähig zu sein. Die Clients nehmen eine passive Rolle ein, sie können zwar Dienste von Servern fordern, aber keine Dienste anderen Clients anbieten. Wenn man sich die heutige Leistungsfähigkeit der Clients ansieht, so ist diese Tatsache der Entlastung durch die Server sicherlich zu vernachlässigen. Sie führt uns sogar vielmehr zur Struktur des P2P-Modells, was nachfolgend noch verdeutlicht wird.

Würden die Clients gleichzeitig einen Server auf dem Rechner laufen lassen, so wäre man schon sehr nah an einem P2P-Netzwerk. Dies war zu der Zeit als es noch wenige Nutzer gab und jeder eine statische IP hatte denkbar. Als das Internet wuchs, waren die Anbieter aufgrund der limitierten IPs gezwungen, diese dynamisch beim Einwählen in das Internet zu vergeben. Diese Tatsache verwehrte den Nutzern die Möglichkeit Server laufen zu lassen, da für einen Zugriff auf diese die IP-Adresse vorher bekannt sein muß. Aus diesem Grund wurden die nützlichsten Dienste auf Servern zentralisiert, welche eine feste IP besitzen und von jedem genutzt werden können, der den Domainnamen kennt.

Ein weiter Grund der gegen das gleichzeitige Einrichten von Servern auf Client-Systemen spricht, ist das Vorhandensein einer Firewall. Von diesen werden die meisten privaten Netzwerke geschützt und unterbinden so einen willkürlichen Verbindungsaufbau zu den eingeschlossenen Rechnern.

Aufgrund dieser beiden Einschränkungen, welche verhindern, daß jeder Rechner Dienste anbieten kann, bleiben viele wertvolle Ressourcen (Speicher und Rechenleistung) ungenutzt. Stellt man sich einmal vor, daß zu jeder Zeit nur 10 Millionen 400 MHz Rechner im Internet sind, von denen jeder ein GB ungenutzten Speicher, 1000 Bps ungenutzte Bandbreite und 10 Prozent ungenutzte Rechenleistung besitzt, so kommt man auf 10 Petabyte (10 hoch 15 Byte) freien Speicherplatz, ca 1,25 GBps freier Bandbreite und 40 hoch 8 MHz verschwendeter Rechenleistung.

Diese Zahlen sprechen für die Idee der P2P-Architektur, bei der jeder Rechner einen Mechanismus erhält, um Dienste untereinander anzubieten. Diese Netzwerke basieren dann nicht mehr auf einem zentralisierten Server, welcher Zugang zu Diensten anbietet. Desweiteren laufen diese Netze üblicherweise außerhalb des Domain-Name-Systems (DNS). Wie man in der nachfolgenden Abbildung erkennen kann, gibt es anstelle der zentralisierten Struktur eine stark vernetzte dezentralisierte Architektur.

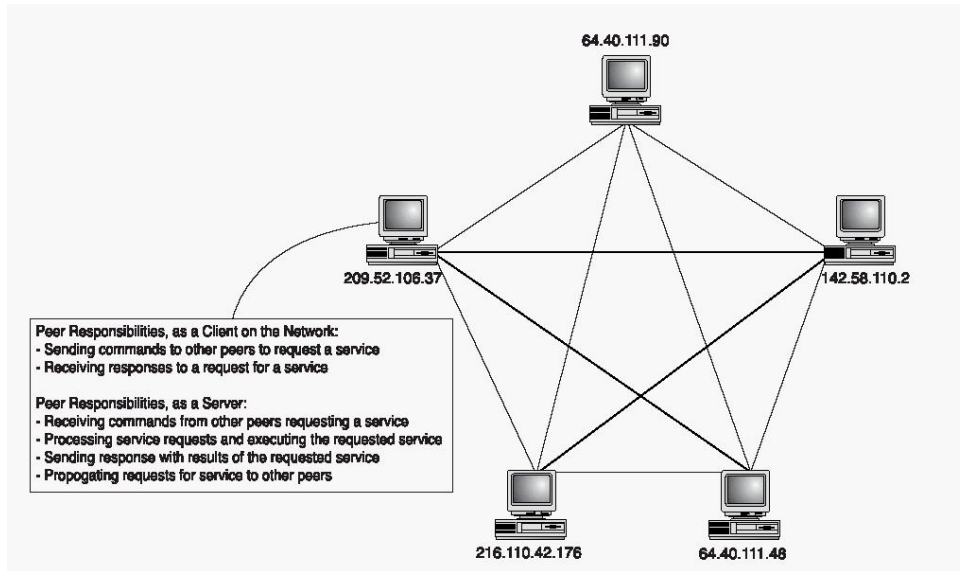


Abbildung 3.3: Peer-to-Peer Architektur [01]

Dadurch, daß bei P2P jedem Rechner die Möglichkeit gegeben wird jeden anderen zu finden, erlaubt man ihnen sowohl als Client, als auch als Server zu agieren. Sie besitzen also die Fähigkeit Dienste anzubieten und angebotene Dienste zu nutzen.

Die Tatsache das die Dienste bei P2P auf alle Rechner verteilt werden, ist ein großer Vorteil, da so Ausfälle von Diensten aufgrund einzelner Fehler ausgeschlossen werden. Durch die dichte Vernetzung, die Vielzahl an Routen und die Verteilung der Dienste wird die verfügbare Bandbreite im gesamten Netz ausgenutzt. Es kommt somit nicht zu den unerwünschten Engpässen, wie man sie im C/S-Modell an den Routen zu viel genutzten Servern vorfindet. Eine Überlastung des Netzes wird dadurch fast ausgeschlossen. P2P offeriert eine viel höhere Verfügbarkeit von Ressourcen zu wesentlich niedrigeren Kosten, durch ein Verteilen der Nachfrage auf das gesamte Netz. Das C/S-System braucht z.B. Bandbreite und Material zu sehr hohen Preisen, um eine ähnliche Robustheit zu erzielen. Viele große Firmen haben deshalb P2P eingeführt, um Kosten bei der Verteilung von Dokumenten und Daten zu sparen. Ein bekanntes Beispiel ist Intel [20]. Ein Einblick in die Kostenverteilung folgt am Ende dieses Kapitels.

Die Redundanz innerhalb der P2P Netzwerke führt auch zu einigen Nachteilen, so sind die vielen möglichen Kommunikationsrouten verantwortlich dafür, daß die Ergebnisse von Dienst Anfragen nicht deterministisch sind. Es kann z.B. vorkommen, daß zwei Clients exakt die gleiche Anfrage stellen, jedoch mit komplett verschiedenen Rechnern über unterschiedliche Wege verbunden werden, welche unterschiedliche Ergebnisse liefern. Es kann auch vorkommen, daß Anfragen sehr lange auf ein Ergebnis warten bzw. kein Ergebnis erhalten. Dies geschieht, wenn ein Rechner der die geforderten Dienste anbietet, einfach vom

Netz geht und läßt sich durch das redundante Vorhandensein dieser Ressourcen umgehen. Dieser Fall ist in der C/S-Architektur unüblich, da die zentralen Server dort üblicherweise ständig erreichbar sind.

Diese Diskussion führt zu folgendem Überblick über die wichtigsten Vor- und Nachteile von P2P:

- **Flexibilität**

Es können jederzeit neue Peers integriert werden, welche dann sofort in der Lage sind, dem Netzwerk ihre Daten und Dienste zur Verfügung zu stellen.

- **Robustheit**

Durch die Redundanz und die dezentrale Struktur ist es möglich, daß einzelne Geräte ausfallen, bzw vom Netz getrennt werden, ohne daß auf Dienste oder Daten verzichtet werden muß.

- **Aktualität**

Wenn Suchergebnisse eintreffen, so kann davon ausgegangen werden, daß diese auch wirklich noch im Netz vorhanden sind, da ein Peer der vom Netz geht, auch nicht mehr auf Suchanfragen antworten kann.

- **Verteilung der Last**

Durch die Verteilung der Dienste über das ganze Netzwerk können, im Gegensatz zur C/S-Architektur, Engpässe vermieden werden, obwohl beide die gleiche Bandbreite benutzen. Nutzer können dabei auf den gleichen Dienst an verschiedenen Stellen zugreifen, und sind so in der Lage, die Netzauslastung gleichmäßig zu verteilen.

- **Sicherheit**

Die Tatsache, daß die Nutzer bei P2P gleichzeitig als Server und als Client agieren können, macht das System wesentlich anfälliger. Während die Server in einem C/S-System von erfahrenen Fachleuten administriert werden, sind die Betreiber bei P2P häufig Laien, welche Sicherheitslücken übersehen.

- **Inkonsistenz**

Ein Nachteil der Redundanz bei P2P ist, daß Daten mehrfach im Netz vorhanden sind. Wird jetzt eine Datei, welche noch auf anderen Peers vorhanden ist verändert, so entsteht eine Inkonsistenz, welche nur durch mühsames Angleichen der anderen Versionen dieser Datei behoben werden kann.

- **Unterschiedliche Standards**

Wie man besonders an den Instant Messengern gesehen hat, führen die unterschiedlichen Standards zu unerwünschten Effekten für die Nutzer. So müssen z.B. alle Messenger-Anwendungen installiert werden, will man an allen Netzen teilnehmen. Unterschiedliche Standards bremsen das Wachstum eines P2P-Netzwerks, da man die Nutzer so unnötig aufteilt. Dies bremst die Entfaltung der Möglichkeiten von P2P.

3.2.2 Beispiel für den Nutzen von P2P

Um zu verdeutlichen, welche Vorteile P2P gegenüber der C/S-Architektur bieten kann, wird nun die Suchmaschine Google [21] näher betrachtet, welche in die C/S-Struktur fällt. Wenn man bei Google eine Suche startet, erhält man üblicherweise als Ergebnis eine Liste mit tausenden von Einträgen, von denen viele ohne Bezug zur Suche, veraltet oder nicht mehr existent sind. Der Grund für diese Problem ist die Zentralisierung der Datenbank. Diese enthält über 1,6 Billionen indizierte Webseiten und ihr Inhalt wird täglich durch Durchsuchen des Internets nach neuen Informationen aktualisiert. Wie man sich denken kann, ist es nicht möglich, daß alle Seiten die in der Datenbank aufgeführt werden auch jeden Tag aktualisiert werden, somit können nicht alle Daten auf dem neuesten Stand sein.

Google weist noch die folgenden Nachteile auf:

- Es benötigt auf der Serverseite viel Ausrüstung. Es benutzt einen Linux-Cluster von 10.000 Rechnern, um den Suchdienst anzubieten.
- Wenn die Suchmaschine vom Netz geht, sind sämtliche Informationen der Datenbank nicht mehr verfügbar.
- Google kann aufgrund der Größe des Internets keine umfassende Tabelle von diesem bieten.
- Die Suchmaschine hat keinen Zugang zu Informationen, die in anderen Datenbanken gespeichert sind.

Betrachtet man jetzt eine P2P-Lösung für eine Suchmaschine wie Google, so muß man sich vorstellen, daß jeder auf seinem Rechner einen persönlichen Webserver laufen lassen könnte. Zu der Fähigkeit, Inhalte von diesem Rechner anzubieten, hätte er noch die Möglichkeit eine Anfrage zu empfangen, diese mit den von ihm verwalteten Dokumenten zu vergleichen und sie mit einem entsprechenden Suchergebnis zu beantworten. Auf diese Art und Weise könnte jeder Rechner viel präziser und aktueller auf Anfragen antworten, welche seine Inhalte betreffen. Es wäre viel leichter die eigenen Dokumente zu indizieren als den Inhalt des gesamten Internets. Unternehmen könnten Gateways anbieten, welche dann interne Datenbanken und deren Informationen für das P2P Netzwerk zugänglich machen. Auch das Problem der Verfügbarkeit läßt sich ausschließen, da ein Rechner der vom Netz geht auch keine Anfragen mehr beantworten kann.

3.2.3 Kosten, Nutzen und Zukunft

Das bei P2P eine auf die Masse verteilte, im Vergleich zu den ansonsten genutzten Servern, günstige Hardware zum Einsatz kommt, wurde bereits erwähnt. Dieses Thema soll nun aufgegriffen und im Detail dargestellt werden. Wie ein einzelner Peer definiert ist, wird in dem Kapitel "Komponenten in einem P2P-System" in "Peer-Typen" beschrieben. Der Gewinn bei P2P liegt in den gesparten Kosten, so geben große Unternehmen wie z.B. Bertelsmann [22] einige Millionen Dollar monatlich aus, um Internetfirmen dafür zu

bezahlen, daß diese für den Konzern den Onlinevertrieb seiner Produkte übernehmen. Allerdings besteht in einem P2P-Netz der Risikofaktor, daß nicht alle für einen bestimmten Dienst erforderlichen User online sind. Die Tatsache, daß Nutzer ihren Computer jederzeit ausschalten können, steht einer hohen Zuverlässigkeit im Weg. Um dieses Risiko zu minimieren müssten Medienunternehmen, welche die P2P-Architektur verwenden, einen eigenen Backup-Server mitlaufen lassen, welcher Daten zur Verfügung stellt, die im Netzwerk gerade nicht verfügbar sind. Dieser Server vermindert die eingesparten Kosten durch seinen Anschaffungspreis. Ein eigener Backup-Server ist langfristig jedoch immer noch billiger als die Beauftragung eines Fremdunternehmens. Bei einem solchen Server fällt ein einmaliger Anschaffungspreis an, welcher bei den Hochleistungsgeräten zwar auch noch im Bereich von einer Million Dollar anzusiedeln ist, allerdings sind die laufenden Kosten (Wartungskosten, Stromverbrauch usw.) wesentlich geringer, als die Rechnungen der Internetfirmen.

Eine weiterer Punkt ist die Frage nach den Nutzungsrechten an den "Privatrechnern". Es wird sicherlich keinen großen Zuspruch geben, wenn die Rechner umsonst ihre Bandbreite und Rechenleistung anbieten sollen. Es wären also zwei Formen denkbar, zum einen eine vertragliche Bindung und zum anderen eine Art Handel.

Bei einem Vertrag müsste man dem Benutzer einen gewissen Betrag für seine Ressourcenabgabe zahlen und dieser würde eine gewisse Verfügbarkeit dieser garantieren. Am sinnvollsten wäre hier sicherlich eine Summe, welche die Anschaffung einer Standleitung großzügig unterstützt und lukrativ werden läßt und so die potentielle Verfügbarkeit der Ressourcen erhöht. Dies bringt dem Benutzer ständigen Zugang zum Internet mit minimalen Kosten und garantiert dem Vertragspartner nahezu ständige Verfügbarkeit.

Ein Handel würde dem Benutzer zwar keine finanzielle Gegenleistung bringen, es wäre aber z.B. denkbar, daß er als Ausgleich einen vergünstigten oder kostenlosen Zugriff auf die offerierten Daten und Dienstleistungen des Unternehmens hat. Dadurch würden zwar die Einnahmen des Unternehmens geringfügig leiden, es gäbe allerdings auch keine zusätzlichen Ausgaben.

Trotz der Vorteile, welche die P2P-Struktur mit sich bringt, gibt es doch mindestens drei Fakten, die gegen eine baldige Ablösung des weit verbreiteten C/S-Systems sprechen. Um die Überlegenheit eines P2P-Netzes zu erreichen und voll auszuschöpfen, bedarf es sehr vieler Peers, die sich beteiligen müssen. Es müßte also ein schneller Umbruch mit großer Beteiligung passieren, um keine inakzeptablen Einbußen zu erleiden. Ein solcher schneller Umbruch ist aufgrund des anfangs mangelnden Anreizes und der fehlenden "Mund zu Mund Propaganda" unwahrscheinlich. Bei einem Server kann die Usergemeinde auch langsam wachsen, ohne daß Nachteile (abgesehen von den Nutzungsgebühren) für den Nutzer entstehen, da hier aufgrund bereitstehender Hardware kein Bedarf an einer großen "Masse" vorhanden ist.

Einem Wechsel auf P2P stehen auch sehr mächtige Interessengruppen gegenüber, welche eine Veränderung bremsen bzw. verhindern werden. Es gibt immer noch eine Menge von Unternehmen, die ihr Geld damit verdienen Inhalte zu vertreiben. Solche erfahrenen Unternehmen mit festem Kundenkreis werden nur schwer aus dem Markt zu drängen sein, zumal die Preise für moderne Server rückläufig sind.

Das Preis/Leistungs-Verhältnis von High-Performance Technical Computern hat sich zwischen 1995 und 2000 um durchschnittlich 37 Prozent pro Jahr verbessert. Dies spiegelt

sich in beeindruckender Weise in den durchschnittlichen Preisen pro GFLOPS wieder. In der 500.000 Dollar Klasse sank der Preis zur Jahrtausendwende innerhalb kurzer Zeit von 14.000 Dollar/GFLOPS auf 6.300 Dollar/GFLOPS. Bei den Geräten über einer Million Dollar sank der Preis im dem selben Zeitraum von 13.000 Dollar/GFLOPS auf 6.400 Dollar/GFLOPS. Dieser Preisverfall wird noch einige Jahre weiter gehen, bevor man schließlich die Grenze des Machbaren erreicht und nur noch kleine Fortschritte in der Entwicklung verzeichnen kann [23].

Zwar werden auch die Rechner der Benutzer billiger und leistungsfähiger, aber ebenso werden die Anwendungen, welche auf ihnen laufen, immer anspruchsvoller und ihre Voraussetzungen wachsen nahezu im selben Tempo. Die ungenutzten Ressourcen auf einem "Privatrechner" können also nicht mit der Entwicklung der Server mithalten.

Wie man gesehen hat, gibt es viele Vorteile die für P2P sprechen, aber auch viele derzeitige Gegebenheiten, welche gegen einen Wechsel sprechen. Es bleibt schwer, eine Prognose zu treffen. Sicherlich steht und fällt P2P mit der Wohlgesonnenheit der Nutzer. Sollte es in diesem Bereich keine große Resonanz geben, so ist ein Durchbruch außerhalb irgendwelcher Tauschbörsen und Messenger Systeme unwahrscheinlich. Allerdings hat man in diesen beiden Bereichen gesehen, wie schnell sich eine gute Idee verbreiten kann.

3.3 Komponenten in einem P2P-System

In diesem Kapitel soll ein grober Überblick über die allgemeinen Bestandteile einer P2P-Architektur gegeben werden. Im Folgenden werden Peers, Peer Groups, Network Transport, Services und Protokolle näher betrachtet.

3.3.1 Peer-Typen

Die Definition von einem Peer gestaltet sich als schwierig, da man seine Möglichkeiten dadurch leicht unvollständig wiedergibt. Allgemein kann man sagen, daß ein Peer ein Knoten in einem P2P-Netzwerk ist, welcher die Basis für die Bearbeitung von Aufgaben in diesem Netzwerk darstellt. Mit der Vorstellung, ein Peer sei eine Anwendung, die auf einem Rechner läuft, welcher an das Netzwerk angeschlossen ist, grenzt man z.B. die Möglichkeiten aus, die eine über mehrere Rechner verteilte Anwendung bieten würde. Ebenso könnte ein Rechner mehrere Peerinstanzen betreiben. Es werden auch Geräte ausgelassen, welche sich indirekt an das Netzwerk anbinden, wie z.B. ein PDA.

In [01] wird die folgende Definition gegeben:

"Any entity capable of performing some useful work and communicating the results of that work to another entity over a network, either directly or indirectly."

Wie man jetzt nützliche Arbeit definiert, hängt davon ab, welche Art von Peer vorliegt. Es werden drei Arten unterschieden, von denen jeder Peer eine oder mehrere repräsentieren

kann. Über die Art bzw. Arten welche ein Peer verkörpert, definieren sich schließlich seine Aufgaben, die er im Netzwerk hat.

Im folgenden werden die verschiedenen Peerarten näher betrachtet:

- **Simple Peers**

Ein Simple Peer hat keine weiteren Aufgaben, als seinem Nutzer die Möglichkeit zu geben seine eigenen Dienste anderen anzubieten und andere Dienste in Anspruch zu nehmen. Aufgrund ihrer eingeschränkten Aufgaben haben sie auch den niedrigsten Grad an Verantwortung im P2P-System. Sie sind im Gegensatz zu den restlichen Peerarten nicht für die Weiterleitung von Nachrichten zuständig.

Im Allgemeinen liegen solche Simple Peers hinter einer Firewall, weshalb sie auch nicht ohne weiteres von außen lokalisiert werden können bzw. mit ihnen direkt kommuniziert werden kann.

Im Grunde ist jeder Rechner in einem P2P-Netzwerk mindestens ein Simple Peer. Allein mit Simple Peers ließe sich z.B. eine interne Datenbank realisieren.

- **Rendezvous Peers**

Ein Rendezvous Peer dient dazu, andere Peers und deren Ressourcen im Netzwerk auf Anfragen hin zu erkunden und die so gewonnene Information dem Interessenten anzubieten. Die Vorgehensweise dieser Peerart, um andere Peers zu erkunden, wird im nächsten Kapitel näher behandelt.

Rendezvous Peers können ihre Fähigkeiten erweitern, indem sie einfach gewonnene Informationen cachen oder Anfragen an andere Rendezvous Peers weiterleiten. Diese Möglichkeiten bieten die entscheidenden Vorteile, daß die Antwortzeiten kürzer werden, umfangreichere Suchergebnisse geliefert werden können und die Netzwerkbelastung reduziert bzw. verteilt wird.

Sinnvollerweise werden diese Peers außerhalb der Firewalls plziert, um die Erreichbarkeit zu gewährleisten. Wenn man sie innerhalb eines von einer Firewall geschützten Privatnetzwerkes unterbringt, muß man ihnen die Möglichkeit geben, diese zu durchdringen. Dies erreicht man entweder mit Hilfe von der Firewall zugelassener Protokolle oder einem Router Peer außerhalb der Firewall.

Rendezvous Peers müssen vorhanden sein, sobald man eine gewisse Suchfunktionalität benötigt, z.B. für eine Suchmaschine, welche mehrere Netze nutzt.

- **Router Peers**

Ein Router Peer ist verantwortlich für die Möglichkeit von außen mit einem Peer kommunizieren zu können, welcher hinter einer Firewall liegt oder durch Network Address Translation vom restlichen Netzwerk getrennt ist. Um eine Nachricht an einen solchen Peer über einen Router zu senden, muß sich der sendende Peer festlegen, welchen Router Peer er nutzen will, um mit dem empfangenden Peer zu kommunizieren.

Mit der Hilfe von Router Peers ist es auch möglich zwischen kommunizierenden Peers zu übersetzen, welche zwei unterschiedliche und inkompatible Netzwerktransporte nutzen.

Router Peers erweitern die Funktionalität in fast allen Anwendungsbereichen dahingehend, daß die Anwendungen auch von privaten Netzen aus genutzt werden können.

3.3.2 Peer Groups

In der Vergangenheit entstanden Unterteilungen innerhalb der Peer-to-Peer Netzwerke, welche aus den unterschiedlichen und inkompatiblen Protokollen der Anwendungen resultieren, die auf einzelnen Peers ausgeführt werden. So können z.B. Peers die Gnutella oder ICQ nutzen auch nur mit Peers kommunizieren, welche ebenfalls Gnutella, bzw. ICQ benutzen. Mittlerweile gibt es P2P-Systeme, bei denen alle Teilnehmer die gleichen Protokolle nutzen können. Das führt zu einem Wegfall der natürlichen Abgrenzungen, wie sie durch frühere Anwendungen entstanden sind und macht eine künstliche Einteilung des Netzwerks in Peer Groups notwendig, um weiterhin Abgrenzungen zu erreichen. In einer Peer Group werden Peers zusammengefasst, um gemeinsamen Zielen und Interessen zu dienen, welche wiederum durch die beteiligten Peers festgelegt werden. Die Dienste die in einer solchen Gruppe angeboten werden können nur von Mitgliedern und nicht von Außenstehenden genutzt werden. Peers einer Peer Group können ihre Dienste redundant anbieten, um sicherzustellen, daß ein Dienst für die Gruppe solange verfügbar ist, solange wenigstens noch ein Peer diesen Dienst anbietet.

Die gemeinsamen Ziele, welche eine Peer Group charakterisieren, basieren auf den folgenden Punkten:

- **Der Anwendung, welche die Peers als Gruppe nutzen wollen**

Dies ist der Fall, wenn die beteiligten Peers Dienste austauschen, bei denen eine Zugriffsmöglichkeit von Außenstehenden unerwünscht ist. Ein Grund dafür wäre z.B., daß die Anwendung geheime oder private Daten nutzt.

Ein Beispiel wäre ein Instant Messenger, welcher nur bestimmten Personen zur Verfügung stehen soll (z.B. interne Kommunikation in einer Firma).

- **Dem Bedarf an Statusinformation von anderen Peers dieser Gruppe**

Mitglieder einer Peer Group sind in der Lage andere Mitglieder zu überwachen. Die so verfügbaren Statusinformationen können benötigt werden, um bestimmte Dienste einer Anwendung zur Verfügung zu stellen, welche von dieser verlangt werden.

Ein denkbare Einsatzgebiet wäre die Bereitstellung von Informationen, über momentan ungenutzte Ressourcen eines Rechners, bei Distributed Computing.

- **Den Sicherheitsanforderungen der beteiligten Peers**

Eine Peer Group ist in der Lage zu entscheiden, wer von außen Zugang zu der Gruppe erhält und Zugriff auf ihre Dienste erlangen darf.

Dies wäre nutzbar, um angebotene Dienste zu verwalten, bzw. verkaufen zu können (Erst nach der Bezahlung wird der Zugang gewährt).

3.3.3 Network Transport

In dem P2P-System JXTA (JXTA steht für "Juxtapose" - Nebeneinanderstellung) [24] dient die Network Transport Schicht der Übermittlung von Daten. Darunter fällt das Aufteilen der Daten in kleine Pakete, welche sich bewältigen lassen und das Anhängen von Paketköpfen, um das Ziel festzulegen. Außerdem besteht die Möglichkeit das Eintreffen

sicherzustellen.

Beim Transport in einem JXTA-P2P-Netzwerk sind drei Elemente beteiligt:

- **Endpoints**

Endpunkte bilden die Quelle oder das Ziel für alle Daten, welche über das Netzwerk versendet werden. Sie können demzufolge Daten senden und empfangen.

- **Pipes**

Pipes sind die Kommunikationskanäle, welche zwei oder mehr Endpunkte miteinander verbinden. Sie können asynchron, unidirektional oder virtuell sein. Die Kommunikation innerhalb einer Pipe verläuft in nur eine Richtung, um eine bidirektionale Kommunikation zu erhalten benötigt man also zwei Pipes zwischen den Endpoints.

- **Messages**

Nachrichten stellen eine Art Kontainer für Daten dar, in denen Daten über die Pipes zwischen den Endpunkten transportiert werden können.

Um mit Hilfe einer Pipe zu kommunizieren, muß ein Peer zunächst seine Endpoints festlegen. Einen als Quelle (Output Pipe) und die oder den Anderen als Ziel (Input Pipe) für die Nachricht. Anschließend werden sie mit einer Pipe verbunden.

Für den eigentlichen Transport der Daten sind nun die Endpoints verantwortlich, da sie selber den Zugriff auf das Netzwerkinterface anbieten. Die Pipe ist lediglich eine Abstraktion, die verdeutlichen soll, das zwei Endpunkte miteinander verbunden sind.

Um nun einen Datentransfer zu vollziehen, werden die Daten, welche transportiert werden sollen von dem Peer in einer Message verpackt, welche anschließend unter Verwendung der Output Pipe versendet wird. Auf der Empfängerseite empfängt ein Peer die Message von einer Input Pipe und entpackt die übertragenen Daten.

3.3.4 Services

Dienste bieten Funktionalität, welche von Peers genutzt werden kann, um Arbeit zu verrichten. Diese Funktionalitäten geben den Anreiz um Geräte an ein P2P-Netz anzuschließen. Ohne Dienste hat man kein P2P-Netzwerk, es existiert dann lediglich eine Ansammlung von Geräten, welche nicht in der Lage sind untereinander ihre Ressourcen zu nutzen. Man kann zwischen zwei Kategorien von Diensten unterscheiden:

- **Peer Services**

Die Funktionalitäten werden von einem bestimmten Peer im Netzwerk anderen Peers angeboten. Die Fähigkeiten des Dienstes sind an den Peer gebunden und verschwinden beim Abmelden mit ihm vom Netzwerk. Sie sind also nur verfügbar, wenn auch der Peer an das Netz angeschlossen ist.

Diese Dienste werden sicherlich eher von Privatpersonen angeboten, welche dem Netz etwas zur Verfügung stellen wollen.

- **Peer Group Services**

Die Funktionalitäten werden hierbei von einer Peer Group an ihre Mitglieder offeriert. Die Fähigkeiten können, wie bereits erwähnt, von mehreren Peers dieser Gruppe angeboten werden und bieten somit redundanten Zugang zu dem Dienst. Solange noch mindestens ein Peer mit diesem Service in seinem Angebot an das Netzwerk angeschlossen ist, so lange existiert auch der Dienst in dieser Gruppe. Dienste in Peer Groups sind meist auf Anbieter, Unternehmen oder Vereine zurückzuführen.

3.3.5 Protokolle

Protokolle sind notwendig, um den Austausch von Daten zu steuern, so legen sie z.B. fest welche Daten übermittelt werden und in welcher Reihenfolge dies geschieht. Sie strukturieren also den Austausch von Informationen zwischen zwei oder mehreren Beteiligten und benutzen zu diesem Zweck Regeln, welche zuvor ausgemacht wurden und allen Beteiligten bekannt sind.

In einem P2P-Netzwerk besitzen Peers die Fähigkeiten für die folgenden Aktionen, welche alle einer Regelung durch Protokolle unterliegen:

- Angebotene Dienste eines Peers finden
- Peers im Netzwerk finden
- Nachrichten für andere Peers routen
- Datenverbindungen zu Peers erstellen
- Statusinformation von einem Peer beziehen
- Dienste eines Peers aufrufen
- Peer Groups erstellen, beitreten und verlassen

3.4 Kommunikation im P2P-Netzwerk

In diesem Kapitel sollen die zwei prägenden Fragen bezüglich des Austausches von Diensten zwischen Netzwerkkomponenten geklärt werden.

Dazu gehört einmal die Frage, wie ein Gerät Peers in einem P2P-Netzwerk findet. Die Bedeutung dieser Frage wird einem schnell klar, wenn man sich überlegt, daß ein Gerät, das nicht von der Existenz eines Peers oder eines Dienstes weiß, auch nicht auf diese zugreifen kann.

Desweiteren stellt sich die Frage, wie ein Gerät aus einem privaten Netzwerk an dem gesamten Netzwerk teilnehmen kann. Die Beantwortung dieser Frage löst das Problem von Geräten, die aufgrund ihrer Positionierung in verschiedenen privaten Netzwerken daran gehindert werden, an dem gesamten Netzwerk teilzunehmen.

3.4.1 Suche nach Peers und Services

Die Bestandteile eines P2P-Netzwerks, wie z.B. Peers, Peer Groups, Services, Pipes und Endpoints lassen sich unter dem Oberbegriff Advertisement zusammenfassen. Dies vereinfacht das Problem der Suche dahingehend, daß man jetzt die allgemeine Suche nach Advertisements im Netzwerk betrachten kann.

Es gibt drei Verschiedenen Möglichkeiten Advertisement in einem Netzwerk zu entdecken, welche im Anschluß aufgeführt werden.

- **No Discovery**

Bei dieser Methode entfällt die aktive Suche nach Advertisements, stattdessen kann ein Peer auf einen Cache mit bisher gefundenen Advertisements zurückgreifen, der Informationen über Peer-Ressourcen bereit stellt. Auf diese Art lässt sich eine Menge Netzwerkverkehr vermeiden und man erhält die Suchergebnisse sofort, im Gegensatz zur aktiven Suche.

Es gibt jedoch den Nachteil, daß auch nicht mehr vorhandene Advertisements im Cache aufgelistet sind, was dem bereits genannten Nachteil des Google-Modells nahe kommt. Der unnötige, fehlgeschlagene Zugriff auf das nicht mehr existente Advertisement erhöht die Netzbelastung wieder. Mit Strategien, welche ein regelmäßiges Überprüfen bzw. Entfernen der Cache Informationen bewirken, lässt sich dieses Problem reduzieren.

- **Direct Discovery**

Diese Methode existiert nur bei Peers, welche sich im gleichen LAN befinden und somit die Möglichkeit besitzen sich direkt zu finden, ohne einen Rendezvous Peer für die Suche hinzu ziehen zu müssen. Für Direct Discovery benutzen Peers die Broadcast- oder Multicastfähigkeiten ihres Network Transport.

Um umliegende Peers zu finden, sendet ein Peer einen Broadcast zu allen Peers im lokalen Netzwerk und erhält daraufhin von allen eine Antwort, mit Informationen über deren Adresse im Netzwerk.

Sobald die Adressen bekannt sind, kann der Peer die anderen Advertisements durch direkte Kommunikation mit den Peers entdecken und ist nicht mehr auf Broadcast oder Multicast angewiesen.

- **Indirect Discovery**

Bei der indirekten Methode des Suchens werden Rendezvous Peers mit einbezogen, um die gewünschten Advertisements zu entdecken. Diese Möglichkeit bietet sich für Peers in einem LAN, wodurch sie nicht mehr auf Broadcast und Multicast angewiesen sind. Ebenso bietet ein Rendezvous Peer auch Peers in einem privaten, von einer Firewall geschützten Netzwerk die Möglichkeit außerhalb liegende Peers zu entdecken.

Rendezvous Peers können bei den Anfragen entweder auch auf gecachte Advertisements zurückgreifen und diese als Antwort nutzen, oder die Anfragen an andere Peers oder Rendezvous Peers weiterleiten. Das Weiterleiten birgt die Gefahr einer hohen Netzauslastung, welche durch ein im Kreis Weitersenden der Anfragen zustande kommt und in der folgenden Abbildung dargestellt wird.

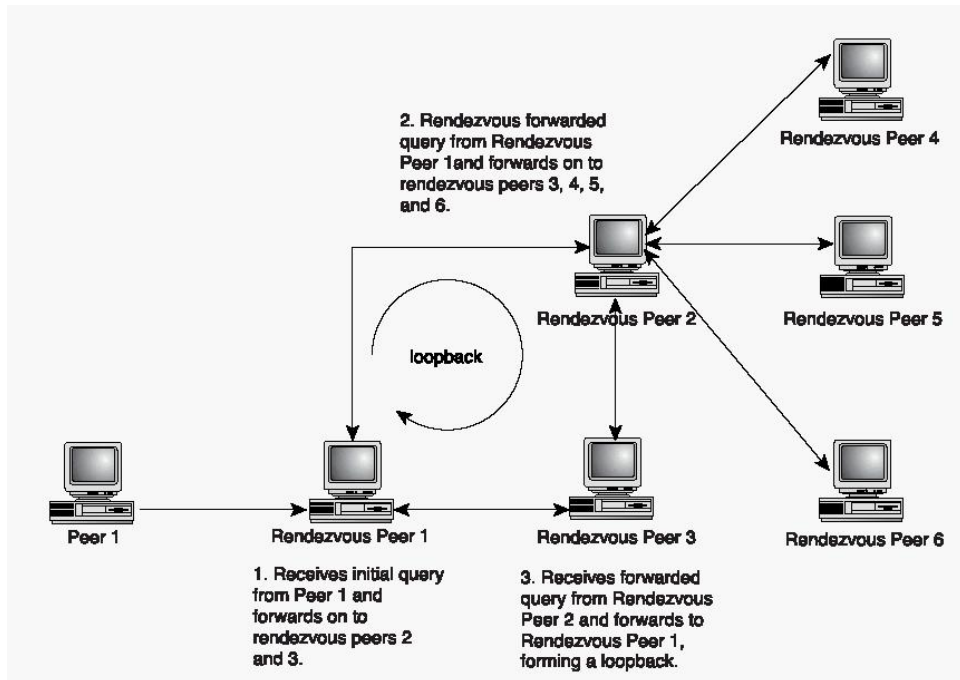


Abbildung 3.4: Loopback [01]

Dieser Effekt namens Loopback lässt sich mit einer festgelegten Lebenszeit der Anfrage entschärfen. Damit legt man die maximale Zahl der Weiterleitungen fest. Kombiniert man diese beiden Wege, so reduziert man die Antwortzeit, verringert die Netzbelastung und hält die Ergebnisse aktuell.

3.4.2 Suche nach Rendezvous und Router Peers

Für Peers in einem privaten Netzwerk gestaltet sich die Suche nach Rendezvous und Router Peers als schwierig, da es aufgrund der Firewall keine Möglichkeit der direkten Suche gibt. Um dem zu begegnen kann man Rendezvous oder Router Peers in das private Netz integrieren und so die indirekte Suche nach außen ermöglichen. Eine einfachere Lösung ist es, den Peers Adressen von festen Rendezvous und Router Peers außerhalb der Firewall vorzugeben und so eine Art Zugangspunkt zum P2P-Netz zu bieten.

3.4.3 Überwinden der Grenzen privater Netzwerke

Um die Frage bezüglich der Teilnahme von Peers eines privaten Netzwerks an dem gesamten P2P-Netz zu beantworten, muß man sie zunächst in drei Fälle unterteilen. Es existieren die Möglichkeiten, daß private Netzwerke durch eine Firewall, Network Address Translation oder eine hybride Lösung von beiden geschützt werden.

- **Firewall**

Eine Firewall ist dazu da, um das private Netzwerk vor unerlaubtem Zugriff zu schützen, bzw. unerwünschte Zugriffe nach außen zu unterbinden. Dieser Mechanismus erschwert allerdings auch die erwünschte Kommunikation mit dem hinter der Firewall liegenden Netzwerk, da üblicherweise nur bestimmte Protokolle bzw. bestimmte Ports freigegeben sind. Um eine Verbindung trotz Firewall herzustellen, muß man also diese Protokolle benutzen und die freigegebenen Ports verwenden.

- **Network Adress Translation (NAT)**

Mit NAT werden IP Adressen eines internen Netzwerkes auf einen Satz von IP Adressen im öffentlichen Netzwerk abgebildet. Hierfür gibt es einmal die statische Variante, bei der für jede interne genau eine externe IP Adresse zur Verfügung steht. Bei der dynamischen Variante existieren mehr interne als externe IP Adressen.

NAT bringt den Vorteil, daß in einem privatem Netzwerk nicht jeder Rechner mit einer globalen IP ausgestattet sein muß, was wiederum Kosten spart. Außerdem lässt sich durch NAT die Sicherheit erhöhen, indem man einen einzigen Zugangspunkt zu dem Netzwerk zulässt. Erreicht wird ein solcher Zugang durch die Regelung, daß nur eingehende Verbindungen zu Rechnern zugelassen werden, welche zuvor eine Verbindung nach außen eingeleitet haben. Somit werden alle Verbindungen zugelassen, welche ihren Ursprung im internen Netzwerk haben.

Die Abbildung von internen IPs auf externe IPs läuft im Router folgendermaßen ab:

- Die Quell IP und der Port des Pakets werden in der Übersetzungstabelle gespeichert
- Die Quell IP des Pakets wird durch eine der externen IPs ersetzt und die Zuordnung wird auch in der Übersetzungstabelle gespeichert
- Der Quell Port des Pakets wird durch einen anderen Port ersetzt und die Zuordnung wird ebenfalls in der Übersetzungstabelle gespeichert

Nach diesen Schritten wurde das Paket weitergeleitet. Kommt ein Paket von außen an einer der vergebenen externen IPs an, so wird der Prozeß umgekehrt durchlaufen und das Paket mit Hilfe der Übersetzungstabelle an die interne IP weitergeleitet. Kommt ein Paket am Router an, welches zu keiner der vergebenen externen IPs gehört, so wird der Zugang verweigert.

- **Firewall und NAT**

Kombiniert man beide Mechanismen, so entstehen weitere Hürden für eine Kommunikation über die Grenze des privaten Netzes hinweg. Wie bereits erwähnt müssen bei einer NAT Verbindungen von außen durch einen Peer innerhalb des geschützten Netzes initialisiert werden und generell können Verbindungen an einer Firewall aufgrund ihrer Quell IP, benutzten Ports oder verwendeten Protokollen blockiert werden. Akzeptiert die Firewall keine Verbindungen nach außen, so ist natürlich keine Kommunikation möglich.

Um eine Verbindung herzustellen, müssen die internen Rechner ein Protokoll verwenden, welches die Firewall passieren kann und bilden so die Initialisierung für eine

Verbindung von außen.

Was bisher unberücksichtigt blieb, ist die Notwendigkeit, daß auch Peers von außerhalb des Privatnetzes in der Lage sein sollten, eine Verbindung zu einem Peer innerhalb der NAT herstellen zu können. Für diesen Zweck wird ein Router vor der Firewall genutzt, zu welchem die externen Peers ihre Verbindung aufbauen, wenn sie eine Verbindung zu einem internen Peer wollen. Die internen Peers müssen sich periodisch mit diesem Router verbinden, um eine Verbindung nach innen zu initiieren. Der Router sendet dann mit dieser Verbindung die vorher empfangenen Nachrichten an den eigentlichen Empfänger weiter. Der Verlauf wird in der folgenden Abbildung illustriert.

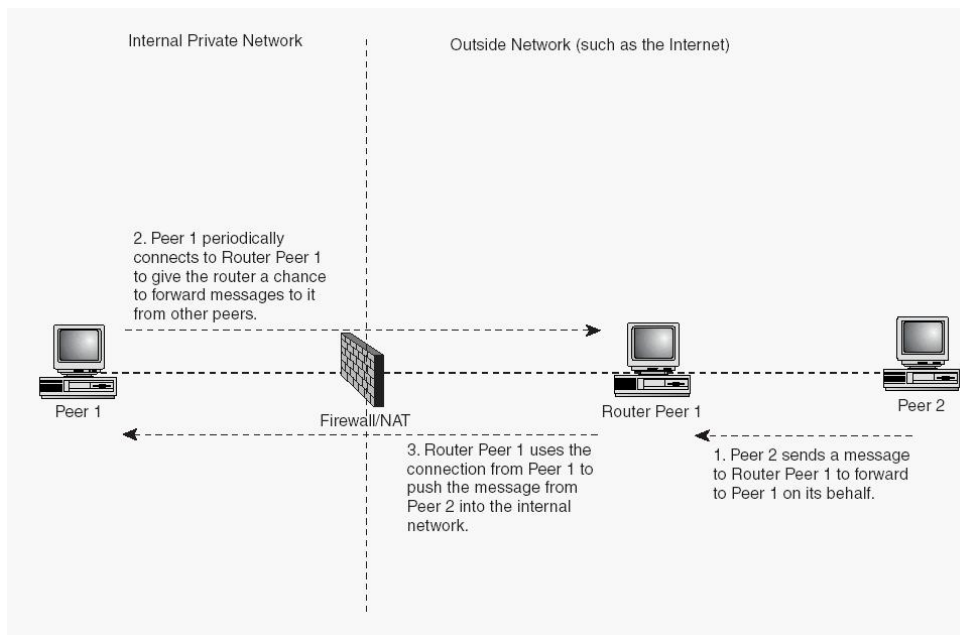


Abbildung 3.5: Hybride Firewall und NAT Lösung [01]

Sollten Sender und Empfänger in privaten Netzwerken liegen, so ist auf der Senderseite auch ein Router zwischengeschaltet und die Abfolge ist analog.

3.4.4 Vorteile der P2P-Kommunikation

Vergleicht man die P2P-Kommunikation mit der herkömmlichen C/S-Kommunikation, so ergeben sich klare Vorteile. Für den Nutzer dürfte besonders erfreulich sein, daß es zu einer Verteilung der Belastung auf das gesamte Netzwerk kommt. Durch die Ausnutzung der gesamten Ressourcen wird eine Überlastung bestimmter Stellen nahezu ausgeschlossen. Dadurch und durch das redundante Angebot von Diensten und Daten werden lange Wartezeiten vermieden, so können z.B. Rendezvous Peers bei Überlastung die Suchanfragen einfach an andere Rendezvous Peers weiterleiten.

Das Angebot der verschiedenen Suchmöglichkeiten gibt den Peers die Flexibilität, die günstigste (schnellste und erfolgreichste) Suchmethode in der bestehenden Umgebung zu wählen. Sollte eine Möglichkeit aufgrund äußerer Umstände nicht möglich sein, so läßt

sich eine andere nutzen. P2P läßt sich außerdem in nahezu jedem Netzwerk verwenden, sowohl mit Firewall als auch mit NAT.

Kommt es zu Änderungen in der Infrastruktur des Netzwerks, so ist P2P aufgrund seiner Kommunikationsmethoden in der Lage, jederzeit kostengünstig an die neuen Verhältnisse angepaßt zu werden. Es werden meist nur neue Leitungen benötigt. Man kann z.B. von Direct Discovery auf Indirect Discovery wechseln, sollte das Netz durch eine Firewall abgegrenzt werden. Neue Peers können einfach in das bestehende Netz integriert werden, ohne ihnen Information über umliegende Peers und deren Dienste mitzugeben.

3.5 Zusammenfassung und Ausblick

Wie man in dieser Arbeit sehen konnte, bietet die P2P-Architektur eine große Anzahl an Vorteilen gegenüber der herkömmlichen C/S-Architektur. Besonders die Flexibilität, die Robustheit und die kostengünstige Realisierbarkeit machen die Attraktivität von P2P aus. Gerade in einer Zeit, in der man für jede Dienstleistung teuer bezahlen muß und in der alles ständigen Veränderungen unterliegt sind das Fakten, vor denen sich niemand verschließen kann. Es kommt sicherlich darauf an, wie dieser wachsende P2P-Markt von Unternehmen vermarktet wird und ob es ihnen gelingt, einen effizienten Umbruch zu verwirklichen, schließlich steht und fällt die Leistungsfähigkeit mit der Teilnehmerzahl. Sollte es nicht gelingen, einen Anreiz für die Abgabe von Ressourcen zu schaffen, so könnte das große Interesse an P2P verschwinden, bevor seine Leistungsfähigkeit erreicht wird.

Sicherlich spielt auch die persönliche Einstellung der Nutzer eine große Rolle, so wird es vielen widerstreben, Ressourcen seines eigenen Rechners zur Verfügung zu stellen. An Freigaben von Daten hat man sich bereits gewöhnt, aber wie sieht es mit Abgaben von Rechenleistung und Bandbreite aus. Hier kommt der Punkt zum tragen, was die Unternehmen den Nutzern als Gegenleistung bieten und ob diese über mögliche Bedenken hinweg hilft.

Ist die schwere Phase der Interessentengewinnung überwunden und hat man genügend Teilnehmer, wird es nur noch wenige Bereiche geben, in denen P2P dem C/S-Modell den Vortritt lassen muß. Zu diesen Bereichen gehören alle Anwendungen, in denen viele Nutzer auf die gleichen Daten zur selben Zeit zugreifen müssen, wie z.B. Flugreservierungen und Aktienmärkte. Hier kommt es zu Inkonsistenz, welche nicht tolerierbar ist.

Die Entwicklung der Abrechnungsverfahren und Kosten für bestimmte Dienste in einem P2P-Netzwerk sind ein weiterer offener Punkt, den man nicht vorhersagen kann. Es bleibt abzuwarten was Nutzer für die Bereitstellung ihrer Ressourcen erhalten und was Firmen für ihre Dienste verlangen werden. Bisher läuft alles auf freiwilliger, bzw. kostenloser Basis.

In Zukunft wird es sicher nicht zu einer Entscheidung zwischen P2P und C/S kommen, man wird beide Systeme nebeneinander vorfinden. Welche von beiden die bessere und sinnvollere Architektur ist, hängt von dem jeweiligen Anwendungsbereich und den damit verbundenen Anforderungen und Kosten ab.

Literaturverzeichnis

- [01] Brendon Wilson: JXTA; <http://www.brendonwilson.com/projects/jxta/>, 2002
- [02] IBM Servers, Mainframe: <http://www.ibm.com/servers>
- [03] Band-X Limited: Virtueller Marktplatz für Bandbreiten und Netzwerke; <http://www.band-x.com>
- [04] TAZ Interview mit Jim Mc Coy:
http://www.taz.de/pt/2001/06/02.nf/isText.idx,13.ausg,is_200102
- [05] Marcus Schögel und Claas van Delden: Peer-to-Peer-Konzepte als Herausforderung für die Distribution der Musikindustrie; <http://www.verkauf-aktuell.de/fb0510.htm>
- [06] Susan Breidenbach: Feature: Peer-to-peer potential;
<http://www.nwfusion.com/research/2001/0730feat.html>
- [07] Gene Kan: Next Step for P2P? Open Services;
<http://www.openp2p.com/pub/a/p2p/2001/08/02/openservices.html>
- [08] Erik Möller: Das Netz der Nutzer; <http://www.heise.de/ct/01/06/080/>
- [09] Clay Shirky: What Is P2P... And What Isn't;
<http://www.openp2p.com/pub/a/p2p/2000/11/24/shirky1-whatisp2p.html>
- [10] ICQ: <http://www.icq.com/company/about.html>
- [11] ICQ: <http://www.icq.com>
- [12] SETI@Home: <http://setiathome.berkeley.edu>
- [13] Yahoo! Messenger: <http://www.yahoo.com>
- [14] MSN Messenger: <http://www.msn.com>
- [15] AOL Messenger: <http://www.aol.com>
- [16] Jabber: <http://www.jabber.com>
- [17] Distribution.net: <http://distributed.net>
- [18] Napster: <http://www.napster.com>

- [19] Gnutella: <http://www.gnutelliums.com>
- [20] Intel: <http://www.intel.com>
- [21] Google: <http://www.google.com>
- [22] Bertelsmann: <http://www.bertelsmann.com>
- [23] Earl Joseph II, Ph.D., Christopher G. Willard, Ph.D., Debra Goldfarb: Price/Performance Trends in High-Performance Technical Computers, 2000
- [24] JXTA: <http://www.jxta.org>
- [25] Usenet: <http://www.usenet.com>

Kapitel 4

VoIP - Harmonization of Protocols and Standards

Michael Beise

Voice-over-IP beschreibt die Technik, Sprache über ein IP-basiertes Netz zu versenden. Dabei wird neben dem Internetprotokoll auch eine Vielzahl anderer Festlegungen und Standards verwendet. Diese sind zum einen die grundlegenden Protokolle TCP, UDP und RTP. Zum anderen gibt es eine große Anzahl unterschiedlicher Architekturen und Lösungsansätze, die ihrer eigenen spezifischen Signalisierungs-, Kontroll- und Übertragungsprotokolle bedürfen. Der H.323-Standard ist so eine Architektur und steht dabei mit anderen Lösungsansätzen, wie zum Beispiel SIP, in Konkurrenz. Dadurch kann es zu Interoperabilitätsproblemen zwischen unterschiedlichen Systemen kommen. VoIP bietet viele Vorteile gegenüber der herkömmlichen Telefonie, hat aber auch noch mit entscheidenden technischen Problemen zu kämpfen.

Inhaltsverzeichnis

4.1	Einleitung	77
4.2	Voice-over-IP	78
4.2.1	Entwicklung von VoIP	78
4.2.2	Funktionsweise von VoIP	79
4.2.3	Das Internetprotokoll	80
4.2.4	Vorteile von VoIP	81
4.2.5	Fax-over-IP	82
4.2.6	Probleme bei VoIP	83
4.3	Standardisierung	84
4.3.1	Gremien	84
4.3.2	H.323	84
4.3.3	SIP	92
4.4	H.323, SIP und Andere im Vergleich	97
4.4.1	H.323 vs. SIP	97
4.4.2	H.323-SIP Interoperabilität	98
4.4.3	MGCP, Megaco und Skinny	98
4.5	Zusammenfassung und Schlussfolgerungen	99

4.1 Einleitung

Die Bedeutung des Internets als ein für jeden zugängliches, weltumspannendes Übertragungsmedium hat seit Mitte der 80er Jahre kontinuierlich zugenommen. Viele neue Dienste und Anwendungen bahnten sich ihren Weg und sind aus unserem heutigen Leben kaum noch wegzudenken. Das Abrufen von Web-Seiten, das Versenden von E-Mails, das Herunterladen von Daten auf den eigenen PC und viele andere Anwendungen, gehören inzwischen zum Alltag. Zunehmend werden Dienste wie Radio hören, Videos ansehen oder das Nutzen von Shop-Portalen interessanter für die Nutzer. Viele neue Anwendungen werden in der nächsten Zeit hinzukommen. Seit 1995 arbeiten Fachleute an Verfahren, um das Internet als Alternative zum klassischen Festnetz oder Mobilfunknetz als Kommunikationsmedium zu nutzen. Die Technologie, die dies ermöglicht, wird mit Voice-over-IP oder IP-Telefonie bezeichnet. Durch die Nutzung bestehender Datennetze zur Sprachkommunikation ergibt sich eine Reihe von Vorteilen. Trotzdem hat Voice-over-IP den Durchbruch zur Mainstream-Technologie noch nicht geschafft. Das liegt vor unter anderem daran, dass es noch einige ungelöste Probleme gibt. Es sind jedoch zahlreiche Hersteller am Markt aktiv, da sich die Technik über kurz oder lang durchsetzen wird.[4] Um die Kompatibilität der verschiedenen Architekturen sicherzustellen, haben Gremien wie die International Telecommunications Union (ITU) oder die Internet Engineering Task Force (IETF) Standards entwickelt. Die wichtigsten sind der H.323-Standard und das Session Initiation Protocol. Dennoch sind im Laufe der Zeit viele andere herstellerspezifische proprietäre Lösungen hinzugekommen, was die Interoperabilität der neu entstehenden Systeme natürlich erschwert.

Im Folgenden soll die Technologie von Voice-over-IP vorgestellt werden. Hierbei wird auf die Besonderheiten bei der Sprachübertragung über IP-basierte Netze im Zusammenhang mit den verschiedenen Versionen des Internet-Protokolls eingegangen. Die sich daraus ergebenden Vorteile werden vorgestellt, ebenso die bestehenden technischen Probleme. Außerdem wird eine weitere Möglichkeit einen Dienst eines leitungsvermittelten Netzes auf ein IP-Netz zu übertragen, nämlich Fax-over-IP, kurz dargestellt. In einem zweiten Schritt werden die beiden wichtigsten Standardisierungsansätze, die H.323-Protokollfamilie der ITU und das Session Initiation Protocol der IETF, beschrieben. Hier soll vor allem ein Einblick in die unterschiedlichen Funktionsweisen der beiden Lösungsansätze gewonnen werden. Dazu werden die jeweils beteiligten Komponenten und Protokolle vorgestellt, um dann schließlich das Zusammenspiel beim Verbindungsaufbau verständlich zu machen. Der dritte Teil dieser Arbeit beschäftigt sich mit der Interoperabilität zwischen den beiden vorgestellten Standards. Anschließend werden H.323 und SIP miteinander verglichen und die Vorteile beziehungsweise Nachteile beider herausgestellt. Außerdem werden abschließend weitere Standards aus dem Bereich Voice-over-IP angesprochen, um die Vielzahl der, sich auf dem Markt befindlichen, Lösungsansätze hervorzuheben. Mittelpunkt der Betrachtungen bleiben jedoch H.323 und SIP, da eine weitere Betrachtung den Rahmen dieser Seminararbeit übersteigen würde.

4.2 Voice-over-IP

4.2.1 Entwicklung von VoIP

Die Idee, das Internet zur Sprachkommunikation zu nutzen, stammt von der israelischen Firma VocalTec. Sie brachte 1995 die ersten Lösungen zur IP-Telefonie auf den Markt. Damit war ein gewöhnlicher PC, ausgerüstet mit einer Soundkarte, einem Lautsprecher und einem Mikrofon, in der Lage, über das Internet mit einem anderen PC zu telefonieren.[8] Die Technik war jedoch nur halbduplex und so mussten die Gesprächspartner, ähnlich wie beim CB-Funk, warten, bis der Gegenüber mit dem Sprechen fertig war. Die erzielte Sprachqualität war ebenso nicht zufrieden stellend. Dennoch erhielt die Idee der Sprachkommunikation über das Internet große Aufmerksamkeit und trotz des massiven Widerstands von 100 US-amerikanischen Telefongesellschaften, die ein Verbot von entsprechender IP-Telefonsoftware gefordert hatten, wurden Internet-Telefondienste durch die US-Bundesbehörde für das Fernmeldewesen zugelassen und diesen letztlich damit auch in Europa der Weg geebnet.[9] In der folgenden Zeit beschäftigten sich mehr und mehr Hersteller damit, die bestehenden Probleme zu lösen und somit Voice-over-IP zu einer echten Alternative zu herkömmlichen Sprachnetzen zu formen. Um die Kompatibilität zwischen den neu entstehenden Systemen zu sichern, entwarf die ITU im Jahr 1996 den H.323-Standard. Dieser definiert eine ganze Familie von Standards, die Verbindung, Sicherheit und Codierung für Multimediakommunikation regeln. Durch die Verwendung von Gateways entstanden neben der ursprünglichen PC-zu-PC Idee weitere Anwendungsmöglichkeiten. PC-zu-Telefon-, Telefon-zu-PC- und Telefon-zu-Telefon-Verbindungen wurden möglich. Letztere Möglichkeit war der Anlass für eine neue Providergruppe ITSP (Internet Telephony Service Provider), die über Prepaid Cards das Telefonieren innerhalb des Provider-Netzes über IP ermöglichten.[8] Ein weiterer Meilenstein auf dem Weg zu marktfähigen Voice-over-IP-Produkten war die Entwicklung von Gatekeepern. Gatekeeper unterstützen die IP-Telefonie, in dem sie Verbindungsaufbau und -abbau steuern. Neben der Aufgabe der Adressberechnung können sie die Bandbreitensicherung übernehmen, um somit eine ausreichende Sprachqualität zu erreichen. In den Jahren 1999 und 2000 entstanden die ersten Komplettlösungen.

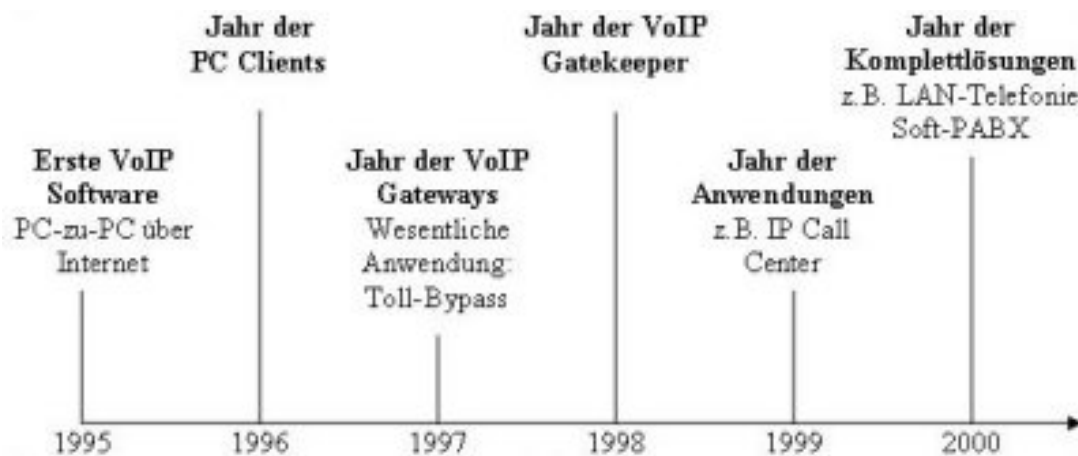


Abbildung 4.1: Meilensteine bei der Entwicklung von VoIP[8]

4.2.2 Funktionsweise von VoIP

Voice-over-IP ist eine Technik der Übertragung von Sprachdaten und Signalisierungsinformationen über Datennetze unter der Verwendung des Internet-Protokolls.[10] Damit besteht der wichtigste Unterschied zur gewöhnlichen Telefonie über leitungsvermittelte Netze darin, dass kodierte Sprachdaten nicht mehr in einem Strom sequentiell übertragen werden, sondern paketierte und auf unterschiedlichen Wegen beim Empfänger ankommen können.

Um die analoge Sprache des Menschen über ein IP-Netz zu übertragen, muss diese, neben der Umwandlung in ein digitales Signal, zunächst kodiert und komprimiert werden. Dazu stehen verschiedene Verfahren zur Verfügung, die, je nachdem wie viel Bandbreite im Netz zur Verfügung steht, genutzt werden können. Soll die Sprache zum Beispiel mit ISDN-Qualität übertragen werden, werden für die Kodierung der Sprache 64 kBit/s benötigt. Um die Qualität eines analogen Telefons zu erreichen, können die Audiosignale auch bis zu einer benötigten Bandbreite von 5,3-16 kBit/s komprimiert werden. Allerdings nimmt mit dem Komprimierungsgrad auch die Zeitverzögerung zu. Dies kann besonders bei VoIP kritisch sein. Anschließend findet die Paketierung der Sprachdaten statt. Gewöhnlich werden bei einer VoIP-Übertragung, neben dem IP-Protokoll, das User Datagram Protocol (UDP) und das Realtime Transport Protocol (RTP) als Transportprotokolle eingesetzt. Legt man eine Kodierung der Sprache mit ISDN-Qualität zu Grunde, wird für eine bidirektionale VoIP-Verbindung, inklusive Overhead durch die Transportprotokolle und Signalisierung, eine Bandbreite von etwa 170 kBit/s benötigt.[8] Nach der Paketierung, gelangen die Sprachpakete auf unterschiedlichen Wegen zum Empfänger. Dabei können sich neben großen Laufzeitverzögerungen auch Veränderungen in der Paketreihenfolge ergeben. Zum Ausgleich wird ein De-Jitter-Puffer verwendet. Sind die Pakete in der richtigen Reihenfolge, können die Daten schließlich dekodiert und ausgegeben werden.

Die Telefonie über ein IP-basiertes Netz kann mit IP-Telefonen oder mit PCs, welche

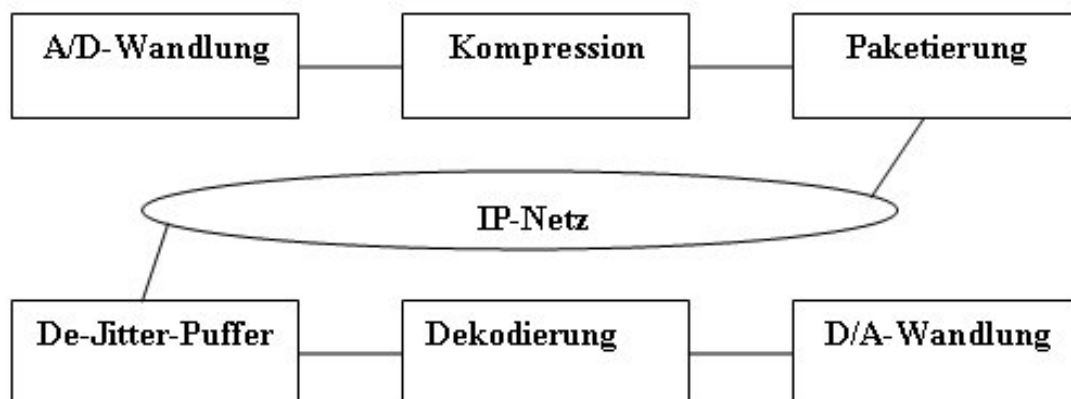


Abbildung 4.2: Funktionsweise im Überblick

über Mikrofon, Lautsprecher, eine voll-dublexfähige Soundkarte und spezieller Telefonie-Software verfügen, realisiert werden. Nutzt der Teilnehmer einen PC, ist er natürlich nur erreichbar, wenn dieser eingeschaltet und die nötige Software geladen ist. Um Gespräche mit Teilnehmern außerhalb des IP-Netzes führen zu können, wird außerdem ein

VoIP-Gateway benötigt. Beim Gateway endet beziehungsweise beginnt die paketorientierte Übertragungstechnik. Die Signalisierung und die zu übertragenden Daten werden hier an das jeweilige Netz des Empfängers angepasst. Durch die Nutzung eines Gateways kann man drei Verbindungsszenarien unterscheiden. Erstens von einem PC zu einem PC innerhalb des gleichen Netzes. Zweitens von einem PC zu einem Telefon in einem anderen Netz oder umgekehrt. Drittens von einem Telefon zu einem Telefon, mit dazwischen liegendem IP-Netz. Die erste Möglichkeit ist relativ einfach. Will ein Teilnehmer eine Verbindung aufbauen, so benötigt er die IP-Adresse der Gegenstelle. Der Verbindungsaufbau, die Verbindungskontrolle und der Austausch der Sprachdaten können dann direkt über die auf den jeweiligen Rechnern installierte Software geschehen. Bei der zweiten Möglichkeit wird das oben beschriebene Gateway benötigt. Will ein Nutzer, der sich in einem IP-Netz befindetet, eine Verbindung aufbauen, so wählt er die IP-Adresse des sich in dem Zielland befindlichen Gateways. Von dort aus können dann die Daten in das lokale Telefonnetz übertragen werden. Beim dritten Verbindungsszenarium wählt der Benutzer die Nummer eines Gateways. Anhand einer zweiten Nummer entscheidet dieses, zu welchem Ziel-Gateway der Verbindungswunsch übertragen werden soll. Am Ziel-Gateway erfolgt nun wieder die Übertragung ins lokale Telefonnetz.[11]

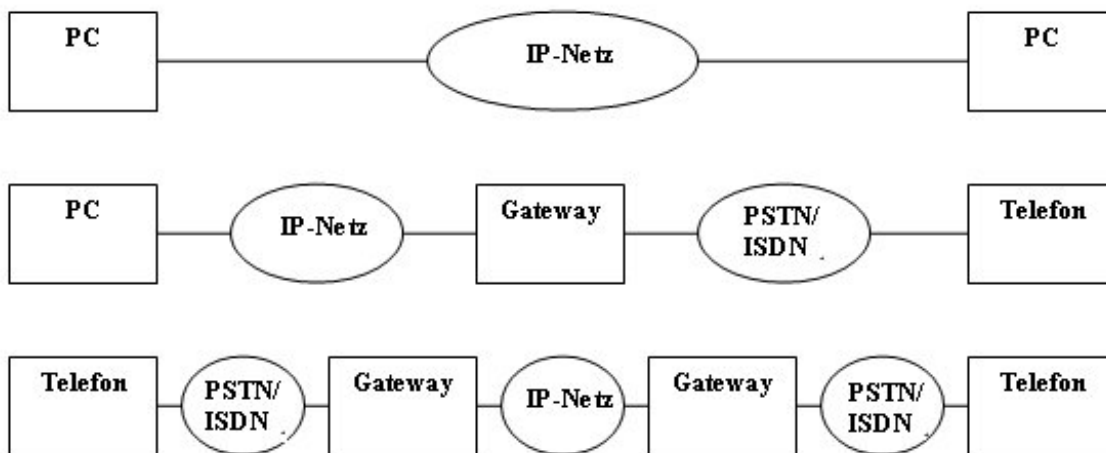


Abbildung 4.3: Anwendungsszenarien

4.2.3 Das Internetprotokoll

Die Nutzung des Internetprotokolls und die darüber liegenden Transportprotokolle sind grundlegend für VoIP, doch ergeben sich durch die Nutzung von IPv4 einige Probleme. Es ist zum Beispiel nicht möglich, auf den Weg, den verschickte Pakete zurücklegen, Einfluss zu nehmen. So können keine Garantien für die Antwortzeit gegeben werden. Im schlimmsten Fall können einzelne Pakete sogar verworfen werden, falls Server oder Router überlastet sind. Man kann zwar erreichen, dass durch das Nutzen des Transmission Control Protocol (TCP) nicht angekommene Pakete erneut versendet werden, jedoch hat das im Bereich von VoIP wenig Sinn, da verspätete Pakete im Kontext einer Sprache

keinen Sinn mehr ergeben würden. Zum Glück ist das menschliche Ohr ziemlich tolerant. Messungen haben ergeben, dass eine Paketverlustrate von fünf Prozent und eine maximale Verzögerung durch die Datenübertragung von bis zu 200 Millisekunden nicht als störend empfunden werden.[10]

Die Verbesserungen in der neuen Version des Internet Protokolls IPv6 kommen auch VoIP zugute. IPv6 bietet eine eingebettete Verschlüsselung und QoS, in dem es dem Sender die Möglichkeit gibt, Datenpakete, die eine spezielle Dienstgüte erfordern, entsprechend zu kennzeichnen. Weiterhin bietet IPv6 eine Vergrößerung des Adressraums, so dass statt bisher 32 Bit, nun 128 Bit Adresslänge vorgegeben werden. Weltweit werden immer mehr Rechner vernetzt, schon allein deshalb ist eine Erweiterung zwingend notwendig, um die Eindeutigkeit der Adressen zu wahren. Zwar konnten mit IPv4 etwa vier Milliarden Adressen dargestellt werden, jedoch sind viele aus unterschiedlichen Gründen nicht verwendbar. Mit IPv6 reicht die Adressvielfalt aus, um jeden Quadratkilometer der Erdoberfläche mit 665.570.793.348.866.943.898.599 Adressen abzudecken.[14] So wäre es auch möglich, mehr statische IP-Adressen zu vergeben, was auch für die Adressierung bei VoIP von Vorteil wäre, da der Nutzer dann immer unter der gleichen IP erreichbar ist. Mit Hilfe von IP-Multicast-Adressen, können auch ganze Gruppen definiert werden, welche bei VoIP-Mehrpunkt-Konferenzen eingeladen werden können.

Das Konzept von Mobile-IP bringt ebenfalls Vorteile für den VoIP-Nutzer. So ist er auch in fremden Netzen über VoIP mit der gleichen Nummer erreichbar.

4.2.4 Vorteile von VoIP

Moderne Unternehmen brauchen zwei Kommunikations- und Informationssäulen, ein Datennetzwerk mit hohem Sicherheitsniveau und schnellen Antwortzeiten sowie ein flexibles Telekommunikations-System. Aus wirtschaftlicher Sicht sind getrennte Infrastrukturen für Datenkommunikation und Telefonie teurer Luxus.[5] Durch Voice-over-IP ist es möglich, beide Säulen über ein Netz zu realisieren. An dieses Netz werden hohe technische Anforderungen gestellt und so sind die Investitionskosten höher als bei normalen Netzen. Der Vorteil liegt jedoch auf der Hand. Neben den geringeren Wartungskosten entfallen die Kosten für Sprachkommunikation innerhalb des eigenen Netzes. Ein einheitliches Management spart Aufwand, Personal und damit Kosten. So hat sich ein solches Universalnetz schnell amortisiert. Durch ein einheitliches Firmennetz können außerdem alle Standorte des Unternehmens die gleiche Kopfnummer besitzen.

Ein weiterer oft genannter Vorteil ist das Einsparen von Telefonkosten bei Ferngesprächen durch die Nutzung des Internets. Hier fallen nur die Kosten bis zum nächsten Einwahlknoten an, also Telefonieren fast überall zum Ortstarif. Besonders interessant ist dieser Faktor für Unternehmen mit vielen weit auseinander liegenden Standorten, sie haben die Leitungen für den gewöhnlichen Datentransport meist sowieso gemietet. Hinzu kommt, dass die Kosten für Sprachkommunikation traditionell in Zeiteinheiten abgerechnet werden, während dies bei Daten zum größten Teil pauschal geschieht. Ist das nicht der Fall, insbesondere bei Privatpersonen, gerät dieser Punkt angesichts der rasant gefallenem Telefongebühren etwas in den Hintergrund.

Durch Kompressionsverfahren lassen sich bestehende Leitungen sehr viel effektiver nutzen, das heißt, es können zum Beispiel auf einem ISDN-Kanal bei Verwendung optimierter Kompressionsalgorithmen mehr als 10 Telefonate gleichzeitig geführt werden. Besonders

für Unternehmen mit bereits bestehenden breitbandigen Intranets ist von einer erheblichen Kostenreduktion auszugehen, vor allem dann, wenn es sich um international tätige Firmen mit zahlreichen Filialen im Ausland handelt.[9]

Mit Voice-over-IP ist es, neben einer sicheren Kommunikation, durch Verschlüsselung möglich, viele Medientypen miteinander zu kombinieren. Nicht zuletzt bieten sich damit viele Einsatzmöglichkeiten für neue Dienste und Anwendungen.

Ein erhebliches Kosten-Einsparungspotential wird insbesondere auch beim Faxdienst gesehen. Etwa 40 Prozent des Volumens im internationalen Telefonverkehr entfallen auf diese Form der Kommunikation, und weltweit werden dafür jährlich ca. 45 Mrd. US-Dollar ausgegeben. Amerikanische Marktbeobachter rechnen damit, dass die Faxkosten mit dem Internet als Transportweg noch stärker fallen werden als die Kosten für IP-Sprachtelefonie.[9]



Abbildung 4.4: Befragung von 2500 Arbeitsstätten[13]

4.2.5 Fax-over-IP

FoIP bietet jedoch noch andere Vorteile. So ist es neben der Variante, Emails als Faxe zu versenden, auch möglich, Faxe als Email direkt an den Emailaccount eines Nutzers zu schicken. Zusätzlich können Faxe mit Passwörtern geschützt oder ganz verschlüsselt werden. Auch die Kombination mit Mobile-IP bringt hier ihre Vorteile, so können Faxe auch in fremden Netzen empfangen werden.

Ähnlich wie bei VoIP, werden Faxe bei FoIP nicht über das Telefonnetz, sondern über ein IP-Netz übertragen. Bei FoIP werden jedoch noch höhere Dienstgüteanforderungen an das übertragende Netz gestellt. Paketverluste können in diesem Bereich nicht akzeptiert werden. Deshalb wird hier TCP, als Transportprotokoll verwendet. Damit können sich noch größere Laufzeitschwankungen ergeben, die wie bei VoIP zu Problemen führen.

Um aus einem IP-Netz heraus Faxe in ein Telefonnetz zu versenden, ist es nötig spezielle FoIP-Gateways einzusetzen. Außerdem lassen sich durch den Einsatz von Gateways zwei Anwendungsvarianten unterscheiden. Die erste Möglichkeit wird mit Store-and-Forward

bezeichnet. In diesem Fall leitet der Sender ein Fax an ein FoIP-Gateway. Dieses speichert das Fax zwischen, übernimmt die Paketierung und leitet es zu einem Ziel-Gateway. Von dort aus wird das Fax zum Endgerät des Empfängers gesendet. Anschließend erfolgt der Versand des Sendebereiches zum Auftraggeber. Da der Ablauf nicht ganz den ganz dem üblichen Faxvorgang entspricht, erfordert er vom Sender ein gewisses Maß an Toleranz.[15] Eine andere Möglichkeit ist das Real-Time-FoIP. Hier wird eine direkte Verbindung zwischen den Endgeräten aufgebaut. Diese Variante entspricht der herkömmlichen Faxübertragung. Der Sender erhält eine direkte Rückmeldung.[15] Das stellt hohe Echtzeitanforderungen an das IP-Netz, welchen man in eigenen LANs durchaus nachkommen kann. Wird aber das Internet genutzt, ergeben sich auch hier viele Probleme.

4.2.6 Probleme bei VoIP

Netze, über die Sprache transportiert werden soll, müssen über ein gutes Echtzeitverhalten verfügen. Das Hauptproblem bei VoIP wurde bereits im Punkt 2.3 angesprochen. Durch die Verwendung des Internetprotokolls kann keine Dienstgüte garantiert werden. Daher müssen Zusatzmechanismen bereitgestellt werden, um den erforderlichen Quality-of-Service zu sichern. Dieser kann zum Beispiel durch Bandbreitenreservierung mittels RSVP (Resource Reservation Protocol), durch geringere Laufzeitschwankungen mittels Multilink-PPP oder durch Priorisierungs- und Queuing-Mechanismen erreicht werden.[12] Innerhalb des eigenen LANs können die bei der Telefonie beteiligten Komponenten so konfiguriert werden, dass keine Bandbreitenengpässe mehr entstehen. Bewegt man sich jedoch in die Welt des Internets, hat man wieder verstärkt mit Laufzeitverzögerungen zu kämpfen, da die Erweiterungen dort nicht immer greifen können. Von der ITU wurde eine maximale Signallaufzeit von 400 Millisekunden empfohlen, da höher liegende Laufzeiten zu einer sehr schlechten Gesprächsqualität führen.[16] Im Internet ist das jedoch meist nicht einhaltbar. Neben der eigentlichen Paketlaufzeit im Netz, kommt die Jitterpufferung hinzu, um unterschiedlich lange Laufzeiten zu kompensieren und die richtige Paketreihenfolge sicher zu stellen, sowie die vom verwendeten Audio-Codec abhängige Zeit für die Kompression beziehungsweise Dekodierung des Sprachsignals. Bei so hohen Laufzeiten muss dann auch eine Echokompensation stattfinden. Weiterhin können ganze Pakete verloren gehen, falls Server überlastet sind. Wenn dann auch keine Interpolationstechniken für eine ausreichende Verbesserung der Gesprächsqualität sorgen können, wird Voice-over-IP für den Nutzer unattraktiv.

Ein anderes Problem im Bereich Voice-over-IP ist die Anzahl verschiedener Protokolle und die Interoperabilität zwischen ihnen. Eines der wichtigsten Konzepte ist der von der ITU entworfene H.323-Standard. Er wurde bei fast allen Herstellern zur Grundlage ihrer Entwicklungen. Da aber H.323 mehrere verschiedene Protokolle umfasst, haben Entwickler die Möglichkeit, spezielle Funktionen hinzuzufügen. Die Folge ist, dass Produkte dann zwar den H.323-Empfehlungen entsprechen, jedoch nicht zwingend mit anderen H.323-Produkten zusammenarbeiten.[6] Der zweite wichtige Lösungsansatz, um Sprache über IP-Netze zu versenden, ist das von der IETF verfasste Session Initiation Protocol. Es ist jünger als H.323 und gilt als echte Alternative. Weitere, in VoIP-Systemen vorkommende Protokolle, sind MGCP, Megaco oder herstellereigene Entwicklungen, wie zum Beispiel Skinny bei Cisco. Will man nun, dass diese unterschiedlichen Systeme zusammenarbeiten, muss viel Geld für Kompatibilitätslösungen aufgewendet werden, die hinfällig werden,

sollte sich ein Standard durchsetzen.

Im Folgenden sollen die Standardisierungsbemühungen bei Voice-over-IP etwas näher beleuchtet werden.

4.3 Standardisierung

4.3.1 Gremien

Es gibt im Wesentlichen vier Gremien, die sich mit der Standardisierung von Voice-over-IP beschäftigen. Diese sind die ITU (International Telecommunications Union), die IETF (Internet Engineering Task Force), das Projekt TIPHON (Telecommunications and Internet Protocol Harmonization over Networks), der ETSI (European Telecommunications Standards Institute) und das IMTC (International Multimedia Telecommunications Consortium).[8] Hierbei ist die ITU das wichtigste Gremium, da es die Normierungskompetenz in der Telekommunikation besitzt. Die ITU veröffentlichte den ersten Standard für Voice over IP, den H.323 Standard. Dieser ist einer der wichtigsten Standards und stützt sich auf viele andere ITU-Festlegungen, wie Audio- und Videocodecs, ab, die bereits aus anderen Telekommunikationsnetzen bekannt sind. Die IETF ist ein internationales Gremium, das sich aus Netzwerkdesignern, -betreibern, Herstellern und Forschern zusammensetzt.[8] Es besitzt keine Normungskompetenz, jedoch sind einige Arbeitsergebnisse, wie das Realtime Transport Protocol von 1996 von der ITU übernommen worden. Die IETF hat auch das Session Initiation Protocol von 1999 ausgearbeitet. Es kann als ernsthafte Konkurrenz zum H.323-Standard der ITU angesehen werden. TIPHON ist ein Projekt, welches sich mit der Interoperabilität zwischen IP-Netzwerken und Circuit Switched Networks, wie PSTN/ISDN, beschäftigt. Die Arbeitsgruppe beschrieb eine Reihe technischer Probleme, welche sich mit der Übertragung von Sprache befassten. TIPHON hat Festlegungen zum QoS getroffen, wie beispielsweise die erlaubte Laufzeit in einem Netzwerk.[8] Die IMTC ist eine Non-profit-Organisation, die den Standardisierungsgremien zuarbeitet. Sie setzt sich aus Herstellern aus der ganzen Welt zusammen.[17]

4.3.2 H.323

H.323 ist ein internationaler ITU-Standard für die Sprach-, Daten- und Videokommunikation über paketorientierte Netze, der die spezifischen Fähigkeiten von Endgeräten im IP-Umfeld festlegt.[18] Dabei legt er die technischen Voraussetzungen für die multimediale Kommunikation in Netzwerken fest, die selbst keinen Quality-of-Service, also Dienstgüte, zur Verfügung stellen.[8] H.323 umfasst dabei, neben Regelungen für das Verbindungsmanagement, also Signalisierung und Verbindungskontrolle, auch eine Beschreibung der für VoIP erforderlichen Komponenten. Die Verarbeitung von Mediaströmen sowie das Interworking verschiedener Netze wurden ebenso durch H.323 standardisiert.[8] Weiterhin wurden Vereinbarungen über Sicherheit und Zusatzdienste getroffen.

H.323 ist eine Rahmenempfehlung, die sich auf etwa 50 andere Standards bezieht. Zu den wichtigsten Festlegungen, gehören H.225.0 und H.245, welche für die Signalisierung

und Verbindung zuständig sind. Für die Codierung der Sprache bezieht sich H.323 auf den Standard G.711, welcher auch bei ISDN seine Anwendung findet. Weitere wichtige Codecs für Sprach- und Videodaten sind G.723.1 und H.261, welche aber nicht bindend in H.323 festgelegt worden sind. Aussagen zur Sicherheit wurden mit H.235 getroffen, Zusatzdienste werden mit H.450.x definiert.

Die erste Version von H.323 wurde 1996 unter dem Titel "Visual Telephone Systems and Equipment for LAN which provide a Non-guaranteed QoS" verabschiedet. Mit der zweiten Version "Packet-based Multimedia Communications Systems" erfolgte 1998 schließlich die Erweiterung auf alle IP-basierten Netze. In dieser und auch in der dritten Version von 1999 kam der H.450-Standard hinzu, um die Leistungsmerkmale von ISDN auch in Voice-over-IP zu integrieren. Er umfasst das Weiterverbinden, Rufweiterleitung, Halten von Anrufen, Anruf parken, Nachrichten übermitteln, Namensübermittlung und Rückruf. Mit der aktuellen vierten Version wurde unter anderem Fax über IP in Echtzeit eingebunden.

Komponenten in einer H.323-Architektur

Eine H.323-Architektur setzt sich aus Terminals, Gateways, Gatekeepern und Multipoint Control Units, kurz MCUs, zusammen. Das Netzwerk selbst ist nicht Teil des H.323-Standards.[8] Terminals, Gateways und MCUs sind Endpunkte in der H.323-Architektur. Zwischen ihnen soll die Kommunikation stattfinden.

Ein Terminal, welches sich in einem LAN befindet, kann zum Beispiel ein IP-Phone oder auch eine Software mit Telefoniefunktion, ein so genanntes Softphone, sein.[8] Sie stellen die Schnittstelle des Benutzers zum System dar und müssen daher mindestens Sprachübertragung, also einen Audio-Codec, unterstützen. Optional ist die Bereitstellung von Video-Codecs. Außerdem müssen alle nach H.323-standardisierten Terminals den Festlegungen nach H.225.0 genügen. Dieser legt fest, wie die Kommunikationsströme an das jeweilige Netz angepasst werden. Dazu gehört auch die Paketisierung. Bei der Signalisierung stützt sich H.225 unter anderem auf die Protokolle Q.931 und RAS ab. Ebenso ist H.245 für alle Terminals verbindlich, um die Zuteilung von Übertragungskanälen und die Vereinbarung über Terminaleigenschaften mit anderen Endgeräten zu ermöglichen.

Gateways in einer H.323-Architektur dienen als Übersetzer zwischen einem IP-Netz und Netzwerken anderen Typs, zum Beispiel ISDN. Auch ist es vorstellbar, klassische Telefone direkt durch Gateways an VoIP-Netzwerken anzuschließen. Derartige Lösungen sind jedoch relativ teuer und nur dann sinnvoll, wenn es sich um hochwertige Endgeräte handelt.[5]

Damit das Gateway seiner Schnittstellenfunktion nachkommen kann, ist es auf der einen Seite mit den Vereinbarungen nach H.323 kompatibel, während es auf der anderen Seite die Übertragungskonzepte und Signalisierungsvereinbarungen eines leitungsvermittelten Netzes unterstützt. Ebenso ist es nötig, die Mediaströme für Sprache und Video in die verschiedenen Codecs zu übersetzen. Die Mechanismen für die Umwandlung der verschiedenen Protokolle, werden im Standard H.246 beschrieben.

Das Gateway ist auch der Netzwerk-/Systemknoten, der für den Jitter-Puffer, die Laufzeitoptimierung, Echo-Unterdrückung und andere Verfahren zur Verbesserung von Quality-of-Service(QoS) zuständig ist.[12]

Da das Gateway in einer H.323-Architektur nicht immer benötigt wird, genau dann, wenn zwei Terminals innerhalb des gleichen Netzes miteinander kommunizieren, ist es eine optio-

Standard	Titel	Version	Publikation
H.323	Visual Telephone Systems and Equipment for LAN which provide a Non-guaranteed QoS	Version 1	November 1996
H.323	Packet-based Multimedia Communications Systems	Version 2	Februar 1998
H.323	Packet-based Multimedia Communications Systems	Version 3	September 1999
H.323	Packet-based Multimedia Communications Systems	Version 4	November 2000
H.323 Annex D	Realtime Fax		September 1998
H.323 Annex E	Framework and wireprotocol for multiplexed call signaling		Mai 1999
H.323 Annex F	Simple Endpoints Type		Mai 1999
H.323 Annex G	Text Conversation and Text SET		Februar 2000
H.323 Annex J	Security for H.323 Annex F		November 2000
H.323 Annex M1	Tunneling of Signaling Protocol Protokoll(QSIG) in H.323		November 2000
H.323 Annex L	Packet-Based Multimedia Communications Systems		März 2001
H.323 Annex M3	Tunneling of DSS1 through H.323		Juli 2001
H.323 Annex Q	Far-end camera control and H.281/H.224		Juli 2001
H.323 Annex R	Robustness Methods for H.323 Entities		Juli 2001
H.323 Annex P	Packet-based multimedia communications systems Annex P: Transfer of modem signals over H.323		Januar 2003
H.225.0	Call Signalling Protocols and Media Stream Packetization	Version 1	Februar 1998
H.225.0	Call Signalling Protocols and Media Stream Packetization	Version 2	September 1999
H.225.0	Call Signalling Protocols and Media Stream Packetization	Version 3	November 2000
H.225 Annex G	Communication between administrative Domains		Mai 1999
H.225 Annex I	H.263 and Video Packetization		September 1998
H.235	Security and encryption for H-Series(H.323) multimedia terminal	Version 1	Februar 1998
H.235	Security and encryption for H-Series(H.323) multimedia terminal	Version 2	November 2000
H.235 Annex F	Hybrid security profile		März 2002
H.245	Media Stream Packetization and Synchronisation on Non-guaranteed QoS LAN	Version 1	März 1996

Abbildung 4.5: Die wichtigsten Standards im Überblick (1)[8]

H.245	Control Protocol for Multimedia Communication	Version 2	September 1998
H.245	Control Protocol for Multimedia Communication	Version 3	Mai 1999
H.245	Control Protocol for Multimedia Communication	Version 4	Februar 2000
H.245	Control Protocol for Multimedia Communication	Version 5	November 2000
H.245	Control Protocol for Multimedia Communication	Version 6	Juli 2001
H.246	Interworking of H-Series multimedia terminals with terminals on GSTN and ISDN		Februar 1998
H.246 Annex C	ISDN User Part Function – H.225.0 Interworking		Februar 2000
H.246 Annex E1	General Inter-Working Function (IWF) between Mobile Application Part and H.225.0		November 2000
H.246 Annex E2	Annex E2: ANSI-41 (Americas) Mobile Application Part and H.225.0 Interworking		November 2000
H.246 Annex F	H.323 - H.324 interworking		Juli 2001
H.261	Video codec for audiovisual services at p x 64 kbit/s		März 1993
H.263	Video coding for low bit rate communication		Februar 1998
H.450.1	Generic functional protocol for the support of supplementary services in H.323		Februar 1998
H.450.2	Call transfer supplementary service for H.323		Februar 1998
H.450.3	Call diversion supplementary service for H.323		Februar 1998
H.450.4	Call hold supplementary service for H.323		Mai 1999
H.450.5	Call park and call pickup supplementary for H.323		Mai 1999
H.450.6	Call waiting supplementary service for H.323		Mai 1999
H.450.7	Message waiting indication supplementary service for H.323		Mai 1999
H.450.8	Name identification services		Mai 1999
H.450.9	Call completion supplementary services for H.323		Februar 2000
H.450.10	Call offering supplementary services for H.323		März 2001
H.450.11	Call intrusion supplementary services		März 2001
H.450.12	Common Information Additional Network Feature for H.323		Juli 2001

Abbildung 4.6: Die wichtigsten Standards im Überblick (2)[8]

nale Komponente. Gateways sind als separate Geräte, PC mit Software oder als integrierte Lösungen, wie zum Beispiel in einem Router, ausgeführt.[8]

Der Gatekeeper ist in einer H.323-Umgebung ebenfalls optional, stellt aber, wenn vorhanden, die wichtigste Komponente dar.[8] Seine Funktionen sind dann auch für alle Komponenten im Netzwerk verbindlich. Ein Gatekeeper erfüllt die zentralen Funktionen, wie Terminal-Registrierung oder Verbindungsauf- und -abbau zwischen registrierten Terminals.[8] Eine Aufgabe des Gatekeepers ist die Signalisierung oder Anrufsteuerung, welche die Umwandlung von IP-Adressen in Telefonnummern beinhaltet. Wie bei den Terminals ist diese Aufgabe in H.225 festgelegt. Die Signalisierung auf den Teilnehmerschlussleitungen zum Auf- und Abbau von Verbindungen mittels Q.931 übernimmt ebenfalls der Gatekeeper. Ein wichtiger Aufgabenbereich des Gatekeepers ist das Bandbreitenmanagement. So überwacht er die Anzahl existierender Gespräche im Netz und lehnt neue Anrufe ab, falls die vorhandene Bandbreite nicht mehr für andere Dienste, wie Email oder File Transfer, ausreichen würde. Alle Terminals, Gateways und MCUs, die ein Gatekeeper verwaltet, werden als Zone bezeichnet. Das Zonenmanagement gehört dann ebenfalls zu den Aufgaben des Gatekeepers, genauso wie die Zugriffskontrolle und das Aushandeln von Leistungsmerkmalen.

Gatekeeper können in Gateways oder Router integriert sein. Das eignet sich aber in der Regel nur für kleine Lösungen bis etwa 30 Terminals.[8]

Um Konferenzen mit mehr als zwei Terminals zu realisieren arbeitet der Gatekeeper eng mit der MCU zusammen. Die MCU besteht aus einem Multipoint-Controller (MC) und einem optionalen Multipoint-Processor (MP). Der MC handelt die zur Verfügung stehenden Ressourcen zwischen allen Teilnehmern, sowohl für Audio- als auch für Video-Daten aus und steuert die Konferenz. Dies ist im Standard H.245 geregelt. Der MP ist für das Multiplexen, Switchen und Verarbeiten von Audio, Video und Daten zuständig.

Eine MCU ist in einer H.323-Umgebung eine optionale Komponente. Sie kann als Stand-alone-Gerät produziert oder in andere Geräte integriert sein.

Verwendete Protokolle

Zur Kommunikation verwendet H.323 sowohl das verbindungsorientierte TCP, als auch das verbindungslose UDP. Zur Signalisierung und Verbindungskontrolle werden bei Voice-over-IP die gesicherten TCP-Kanäle genutzt. TCP sorgt vor allem dafür, dass bei der Übertragung keine Pakete verloren gehen, indem es diese erneut sendet. Hier können sich große Laufzeitschwankungen ergeben. Dies ist unter anderem ein Grund dafür, dass die Audio- und Videoströme selbst, mittels UDP übertragen werden, da zeitversetzt ankommende Pakete sinnlos sind, wenn die Verzögerung so groß ist, dass der De-Jitter-Puffer nicht ausreicht. Ein durch die Verwendung von UDP hervorgerufener Paketverlust, der kleiner als fünf Prozent ist, ist tolerierbar und wird durch den Menschen nicht wahrgenommen. Das UDP sorgt durch seine Verbindungslosigkeit für eine schnellere Übermittlung und zudem bei einem Header von nur acht Byte für weniger Overhead.[8] Allerdings muss hier auf die Flusskontrolle verzichtet werden, welche Bestandteil von TCP ist. Deshalb wird bei der Übertragung von Audio- und Videodaten mittels UDP das Realtime Transport Protocol, kurz RTP, eingesetzt. Das RTP kennzeichnet die transportierten Informationen und fügt einen Zeitstempel sowie eine Sequenznummer ein.[8] Dadurch wird gegenüber Laufzeitschwankungen und Paketverlusten eine gewisse Toleranz geschaffen.

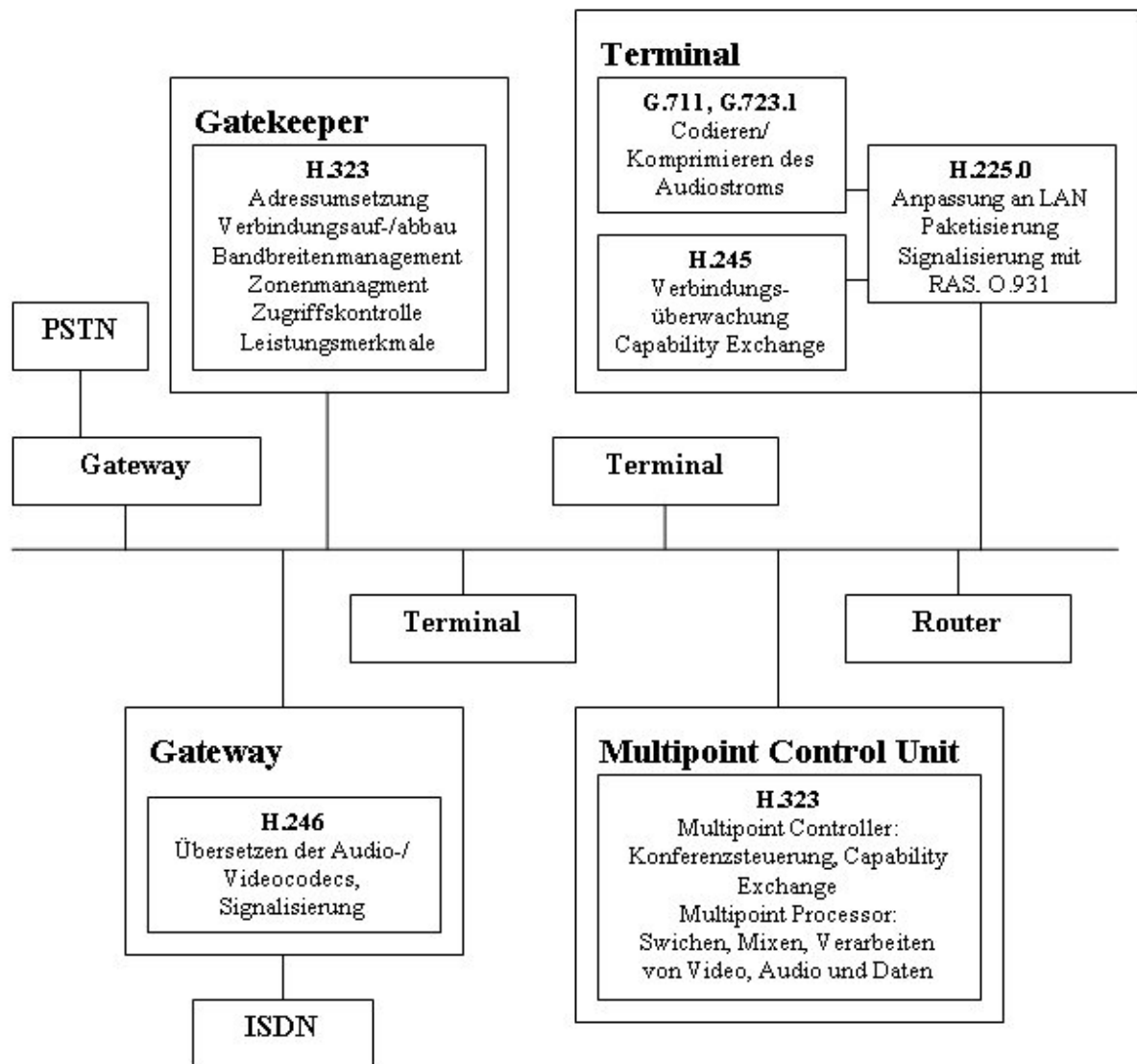


Abbildung 4.7: H.323-Komponenten im Überblick

Kommt es zu Laufzeitschwankungen, so können verspätete Pakete immer noch eingebunden werden. Bei zu großen Laufzeiten oder Paketverlusten müssen beim Empfänger Interpolationstechniken eingesetzt werden. Um die verwendete Bandbreite so niedrig wie möglich zu halten, werden durch Voice Activity Detection Sprachpausen entdeckt. In diesem Fall werden keine Pakete gesendet, jedoch kann durch den von RTP eingefügten Zeitstempel, der Verlust der Isochronität beim Empfänger wieder ausgeglichen werden. Das Realtime Transport Control Protocol, kurz RTCP, ist das Steuerungsprotokoll für RTP. Es setzt ebenfalls auf UDP auf und bezieht sich immer auf eine RTP-Session. Mit Hilfe von RTCP kann der Empfänger Informationen über die Teilnehmer oder über die Verbindungsqualität erhalten.

Die Signalisierung und damit Steuerung einer Verbindung setzt sich aus der RAS-Signalisierung, der Call-Signalisierung und der Call-Control-Signalisierung zusammen. Die Registration, Admisson and Status, RAS-Signalisierung findet zwischen Endgeräten und den Gatekeepern statt. Endgeräte können sich damit registrieren lassen oder einen Verbindungswunsch an den Gatekeeper stellen. Dieser kann dann, zum Beispiel anhand der zur Verfügung stehenden Bandbreite, die Zugangsberechtigung erteilen. Der Verbindungsauf- und -abbau zwischen zwei H.323-Endgeräten fasst sich unter der Call-Signalisierung zusammen. Sie findet mittels Q.931-Messages statt, die auch bei ISDN ihre Anwendung finden.[8] RAS- und Call-Signalisierung sind in H.225 spezifiziert. Durch die Call-Controll-Signalisierung werden Kontrolldaten, wie Kapazitäten oder Terminaleigenschaften zwischen, den Endgeräten ausgetauscht. Sie wurde in H.245 festgeschrieben. Während die RAS-Signalisierung einen UDP-Kanal nutzt, sind Verbindungsaufbau und Verbindungskontrolle von großer Bedeutung und werden über einen gesicherten TCP-Kanal übertragen.

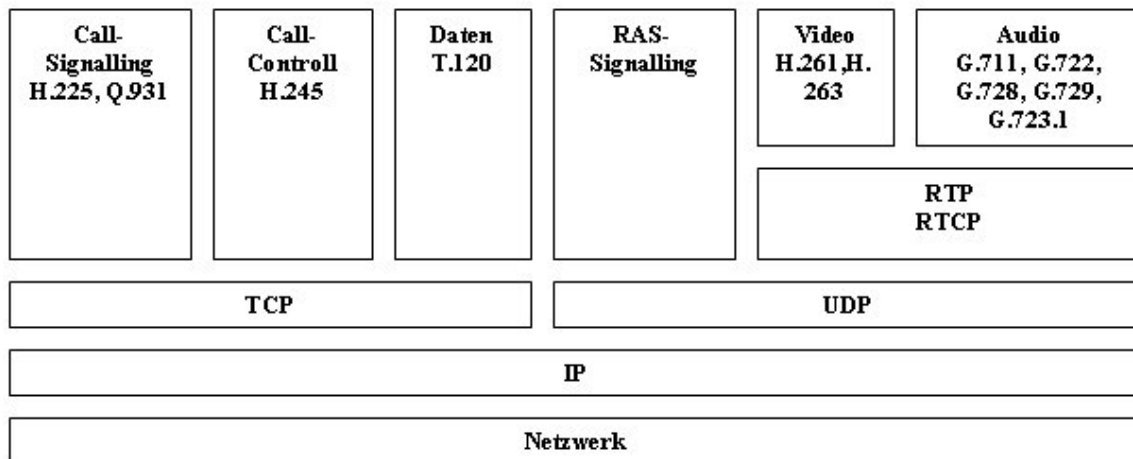


Abbildung 4.8: Verwendete Protokolle im Überblick

Verbindungsaufbau

Innerhalb eines LANs kann der Verbindungsaufbau zwischen zwei Terminals mit oder ohne Gatekeeper erfolgen. Ist dieser jedoch vorhanden, müssen die Teilnehmer vor Gesprächsbeginn ihr Vorhaben dem Gatekeeper durch das RAS-Protokoll signalisieren. Dies

setzt voraus, dass das Terminal auch beim Gatekeeper registriert ist. Nachdem dieser den Verbindungsaufbau bewilligt und die Adresse des anzurufenden Teilnehmers in eine IP-Adresse aufgelöst hat, kann der eigentliche Verbindungsaufbau vollzogen werden. Dazu baut der anrufende Teilnehmer einen TCP-Kanal zum Zielterminal auf. Über diesen Kanal läuft nun die Call-Signalisierung und die Call-Controll-Signalisierung ab. Stimmt der angerufene Teilnehmer der Verbindung zu, so öffnet der Anrufer einen zweiten TCP-Kanal, über den nun die Steuerungsfunktionen ausgetauscht werden. Es folgt der Austausch der Terminalfähigkeiten, das Aushandeln der Rollen, also die Master/Slave-Verhandlung, die Bestimmung der möglichen Betriebsmodi und die Mitteilung über verwendete Audio-/Videocodecs. Der erste TCP-Kanal wird nun nicht mehr benötigt und kann wieder abgebaut werden, dafür wird nun ein RTP- und dazugehöriger RTCP-Kanal geöffnet. Über den RTP-Port erfolgt nun die Übertragung der Sprach- und Videostreams. Befindet sich ein Gatekeeper im Netz, kann die Call-Signalisierung auch über ihn laufen, dies ist optional.

Befinden sich Anrufer und Anzurufender nicht im gleichen Netz, wird ein Gateway be-

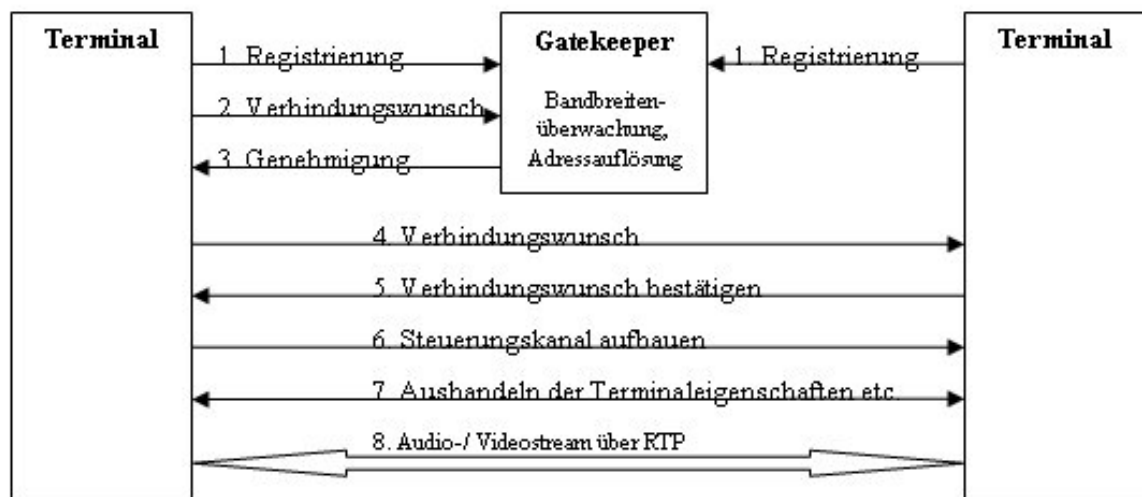


Abbildung 4.9: Verbindungsaufbau Terminal und Terminal

nötigt. Hierzu baut ein Terminal eine Verbindung zum Gatekeeper auf, welcher die IP-Adresse des zu verwendenden Gateways und ihm seinerseits den Verbindungswunsch signalisiert. Das Gateway baut nun die Verbindung zum anzurufenden Teilnehmer auf, in dem es die spezifischen Signalisierungsprotokolle und -prozeduren des Zielnetzes nutzt. Nimmt der Angerufene das Gespräch an, etwa durch Abheben des Hörers, so teilt das Gateway dies dem Gatekeeper mit. Dieser kann nun dem Terminal mitteilen, dass der Verbindungsaufbau und der nachfolgende Audio-/Videostream direkt über das Gateway ablaufen.

Um Multipoint-Konferenzen aufzubauen, wird bei der Verwendung von H.323 eine MCU benötigt. Hier gibt es zwei Möglichkeiten. Eine ist es, eine Pre-arranged-Konferenz aufzubauen. Das bedeutet, dass sich alle Teilnehmer mit einer separaten MCU, die die Konferenz kontrolliert, verbinden. Die andere Technik ist eine ad-hoc-Konferenz. Das bedeutet, dass ein Zweiparteianruf auf drei oder mehr Teilnehmer erweitert wird.[20] In diesem Fall gibt das Master-Terminal die Kontrolle an die MCU ab. Der Multipoint Controller

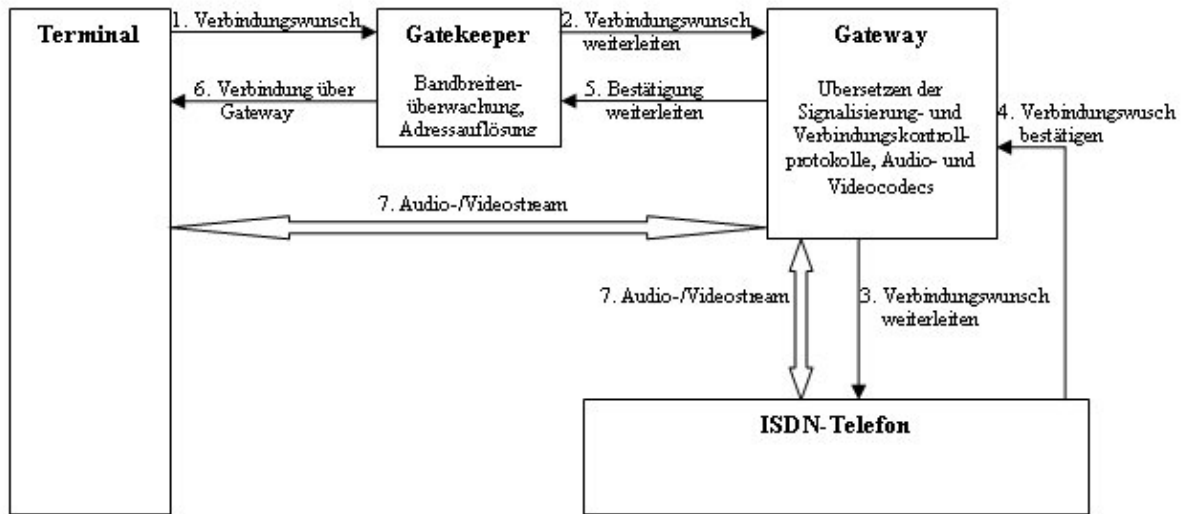


Abbildung 4.10: Verbindungsaufbau Terminal und Telefon

übernimmt in beiden Fällen die Aushandlung der Terminaleigenschaften, sowie die gesamte Steuerung der Konferenz. Ist in der MCU ein Multipoint Processor vorhanden, wird hier sofort das Mixen, Switchen und Verarbeiten aller Datentypen übernommen.

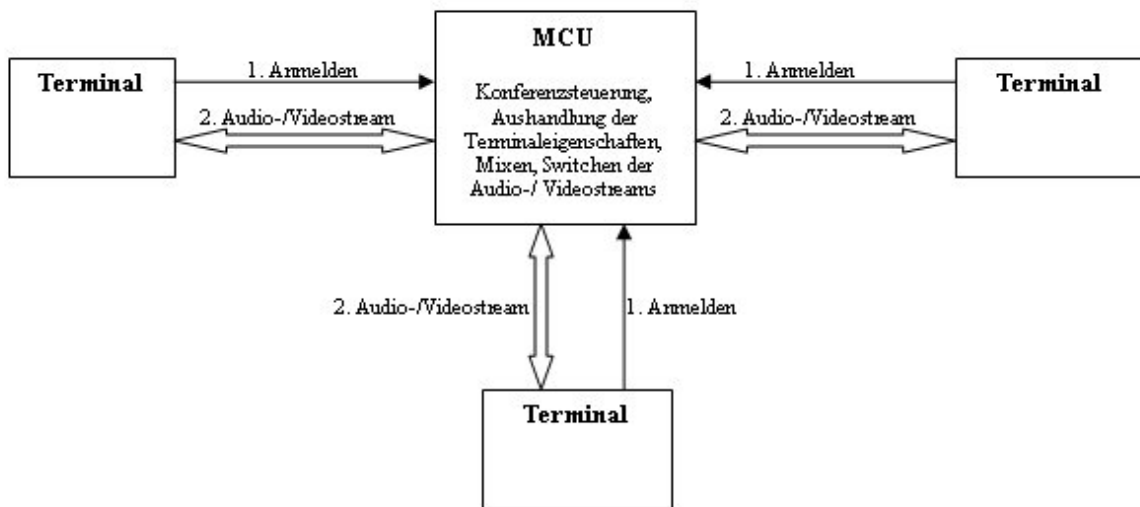


Abbildung 4.11: Verbindungsaufbau Konferenz

4.3.3 SIP

Das Session Initiation Protocol, kurz SIP, wurde 1999 von der IETF unter dem Originalnamen "Application-level protocol for inviting users to multimedia conferences" entworfen

und gilt als ernsthafte Konkurrenz zum H.323-Standard. SIP dient als Signalisierungsprotokoll zum Auf- und Abbau von Multimediasitzungen mit zwei oder mehreren Teilnehmern. Im Gegensatz zu H.323 handelt es sich bei SIP um keine Familie von Standards, es ist nur das Protokoll selbst gemeint. Damit ist es viel einfacher und kann mit anderen Protokollen leicht kombiniert werden. Es wird häufig, in einer von der IETF entworfenen Architektur, zusammen mit dem Session Announcement Protocol und dem Session Description Protocol in einer Client-Server-Umgebung verwendet.

Standard	Version	Publikation
Session Initiation Protocol		März 1999
Session Description Protocol		April 1998
Session Announcement Protocol	Version 1	November 1996
Session Announcement Protocol	Version 2	Oktober 2000
Resource Reservation Protocol		September 1997
Realtime Streaming Protocol		April 1998
Realtime Transport Protocol		Januar 1996

Abbildung 4.12: Die wichtigsten Protokolle im Überblick

Komponenten in einer SIP-Architektur

Wie oben bereits erwähnt, wird in einer SIP-Architektur das Client-Server-Prinzip verwendet. Dabei wird der dienstanfordernde Client auch als User-Agent-Client und der dienstbearbeitende Server als User-Agent-Server bezeichnet. Ein Teilnehmer kann die Eigenschaften eines UAC und eines UAS in sich vereinen, dann wird er als User-Agent bezeichnet. Damit wird mit Hilfe von SIP auch eine peer-to-peer Kommunikation möglich.

Eine SIP-Umgebung kann vier unterschiedliche Servertypen beinhalten. Diese sind SIP-Proxy-Server, Redirect Server, Registrar und Location-Server. Der Proxy-Server bearbeitet Anfragen eines Clients entweder selbst oder leitet diese an einen anderen Server weiter. Beim Zielsystem sieht die weitergeleitete Anfrage so aus, wie eine vom Client kommende Anfrage.[20] Ein Redirect-Server initiiert selbst keine SIP-Sitzungen, er wandelt ankommende Adressen in eine oder mehrere neue Adressen um und schickt diese zurück zum Client. Dieser kann dann eine Anfrage an die neue Adresse stellen.[20] Der Registrar bearbeitet Anmeldungen von Clients im System. Damit kann der Benutzer seine Verfügbarkeit mitteilen. Der Registrar ist üblicherweise mit einem Proxy-Server oder Redirect-Server kombiniert.[20] Ein Location-Server bietet den so genannten Location-Service an, der die Zieladressen von Benutzern angibt und damit eine Weiterleitung bzw. Umleitung eines Anrufs an die richtige Adresse ermöglicht.[20]

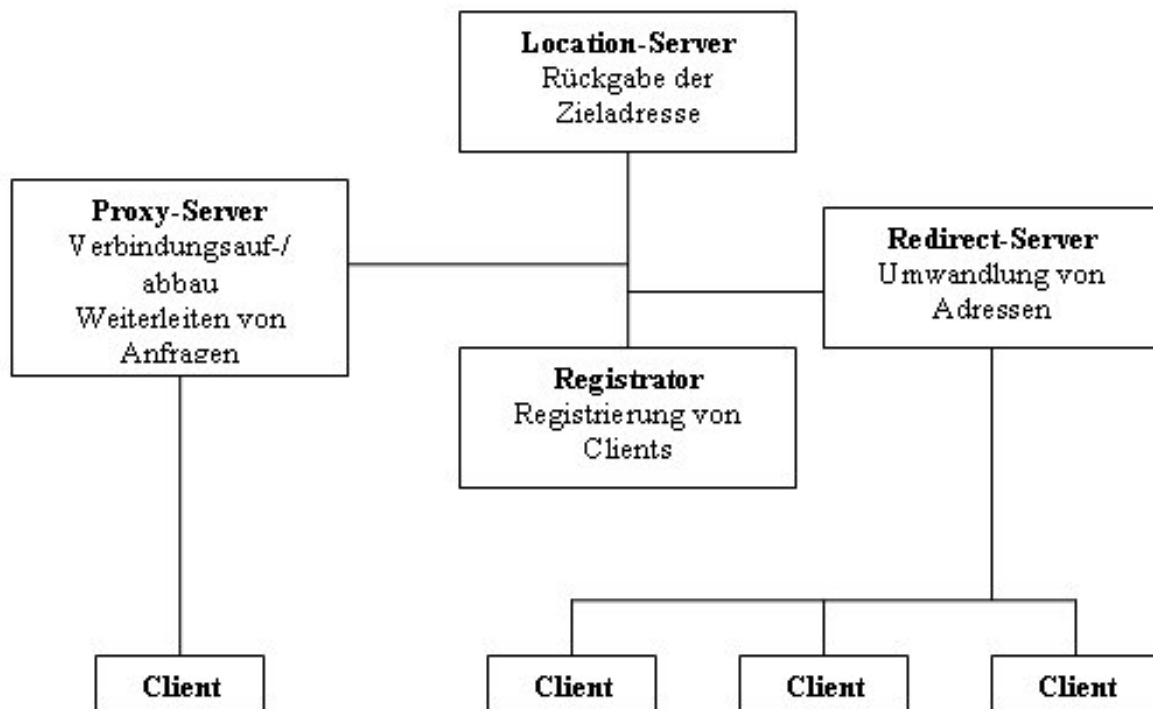


Abbildung 4.13: SIP-Komponenten im Überblick

Verwendete Protokolle

SIP ist Teil einer von der IETF entwickelten Multimedia- und Kontrollarchitektur. Daher arbeitet es häufig eng mit anderen von der IETF entworfenen Standards zusammen, ohne jedoch von ihnen abhängig zu sein. Zur Unterstützung dienen das Session Announcement Protocol (SAP), das Session Description Protocol (SDP), RSVP, das Realtime Streaming Protocol (RTSP) und RTP. SIP arbeitet auf der Anwendungsebene, OSI-Schicht 5 bis 7, und kann daher sowohl TCP als auch UDP als Transportprotokoll nutzen.[8] SAP dient zur Anzeige von Multimediasitzungen über Multicast. Durch SDP erfolgt die Beschreibung einer Multimediasitzung. Hiermit können alle relevanten Informationen für eine Session übermittelt werden. Es beinhaltet Angaben zum Eigentümer, zur Multimediafähigkeit des Nutzers, Bandbreiteninformation und andere. RSVP ist ein Signalierungsprotokoll zur Bandbreitenreservierung. Dabei teilt der Teilnehmer dem Netzwerk seinen Bandbreitenbedarf und die maximale Verzögerungszeit der Datenpakete mit. Alle Kommunikationselemente im möglichen Kommunikationspfad bekommen daraufhin eine Anfrage. Kann die Einhaltung der Bedingungen zugesichert werden, wird die Verbindung durchgeschaltet.[12] RTSP erstellt und kontrolliert die Audio- und Videoströme zwischen den beteiligten Elementen in einer SIP-Umgebung. Zwar kann in einer SIP-Sitzung ein beliebiges Transportprotokoll genutzt werden, jedoch ist RTP das am häufigsten genutzte Protokoll.

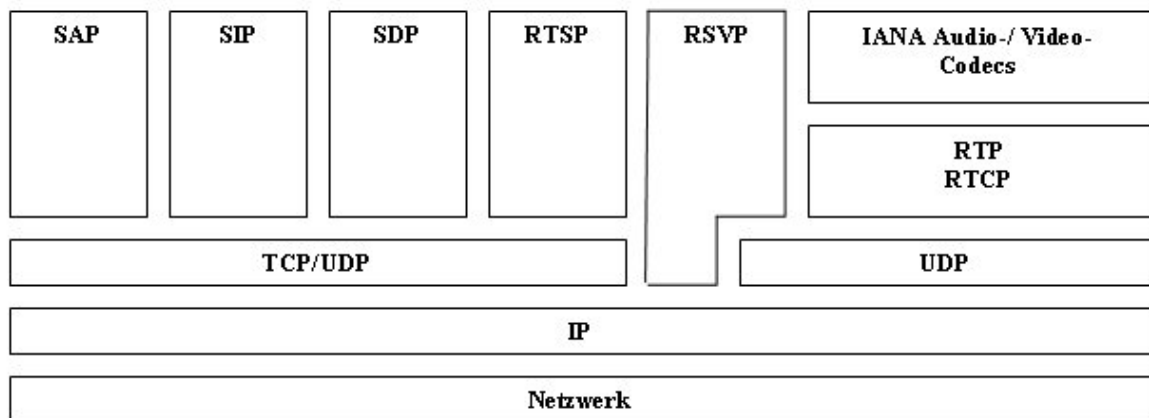


Abbildung 4.14: Verwendete Protokolle im Überblick

Verbindungsaufbau

Nachdem sich ein Client beim Registrator angemeldet hat, kann der Aufbau einer SIP-Session mittels SIP-Proxy-Server oder Redirect-Server erfolgen. Dem kann auch noch eine Bandbreitenreservierung mittels RSVP vorangehen.

Im Proxy-Mode stellt ein User-Agent-Client eine Anfrage an einen Proxy-Server. Dieser reicht die Anfrage an den Location-Server weiter, welcher die Adresse des Anzurufenden ermittelt und an den Proxy-Server zurückgibt. Nun teilt er dem anzurufenden Teilnehmer den Verbindungswunsch des Antragstellers mit. Ist der Verbindungswunsch erfolgreich, teilt der anzurufende Teilnehmer dies dem Proxy-Server mit. Der Proxy leitet die Antwort an den Antragsteller weiter, welcher seinerseits mit der Bestätigung reagiert. Nachdem diese vom Proxy an den anderen Teilnehmer weitergeleitet worden ist, kann mit der Datenübertragung begonnen werden. Diese findet meistens mittels RTP und RSTP statt.

Die zweite Möglichkeit eine SIP-Session aufzubauen bietet der Redirect-Mode. Hier sendet ein Client eine Anfrage an den Redirect-Server. Dieser fordert die Adresse des anzurufenden Teilnehmers vom Location-Server an. Anstatt jedoch einen Verbindungsaufbau zum Anzurufenden zu starten, leitet er die Adresse an den Antragsteller weiter. Dieser bestätigt den Erhalt der Adresse und sendet seinen Verbindungswunsch direkt an den zweiten Teilnehmer. Ist die Verbindung gewünscht, so teilt er dies dem Anrufer mit. Nach einer Bestätigung seinerseits kann die Datenübertragung beginnen.

Um zu anderen Netzen Verbindungen aufzubauen, benötigt man bei SIP selbstverständlich wieder ein Gateway, um die Signalisierung, Verbindungsüberwachung und verwendete Codecs entsprechend anzupassen.

Beim Aufbauen von Multipoint-Konferenzen wird bei SIP keine zentrale Einrichtung benötigt, da jeder SIP-Nutzer selbst Konferenzen unterstützt. Der Aufbau kann auf zwei unterschiedliche Weisen geschehen. Mit Hilfe von SAP ist es möglich Sitzungen anzukündigen. Dies kann durch Email, Webseiten oder Ähnlichem geschehen. Die andere Möglichkeit ist, mit SIP die Teilnehmer über Multicast direkt einzuladen. In beiden Fällen ist auch hier SDP für die Beschreibung der Sitzung verantwortlich. Während der Session werden von den Teilnehmern SAP-Nachrichten gesendet, bleiben diese aus, wird die Konferenz beendet. SAP ist auch für das Aufheben von Konferenzen verantwortlich.

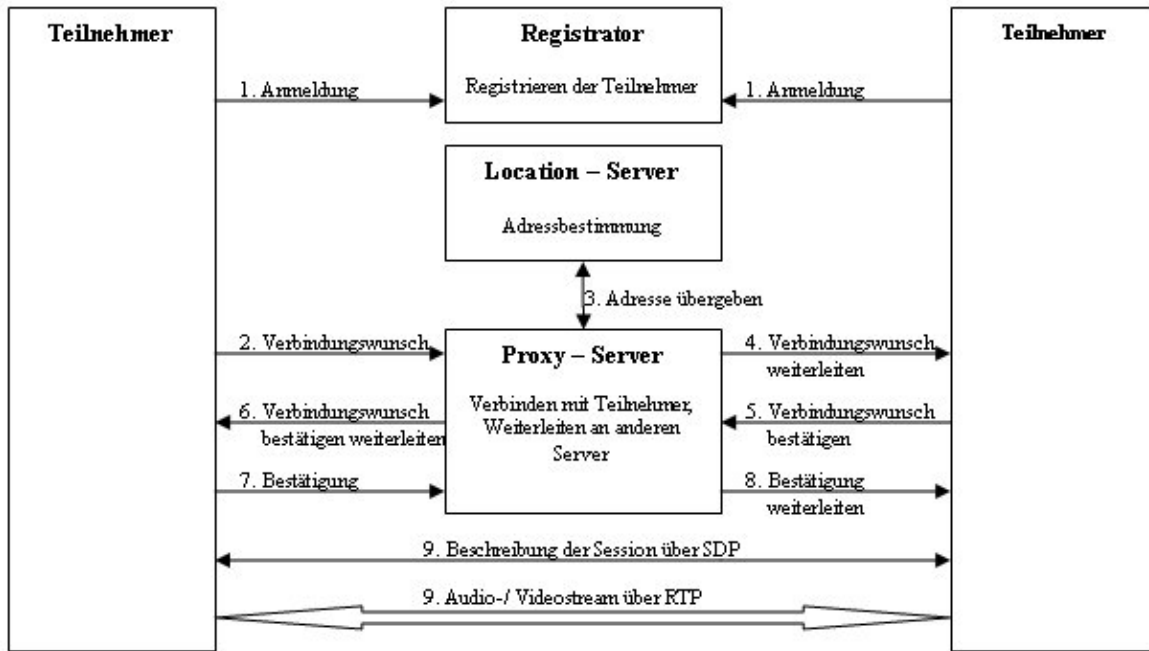


Abbildung 4.15: Verbindungsaufbau über Proxy-Server

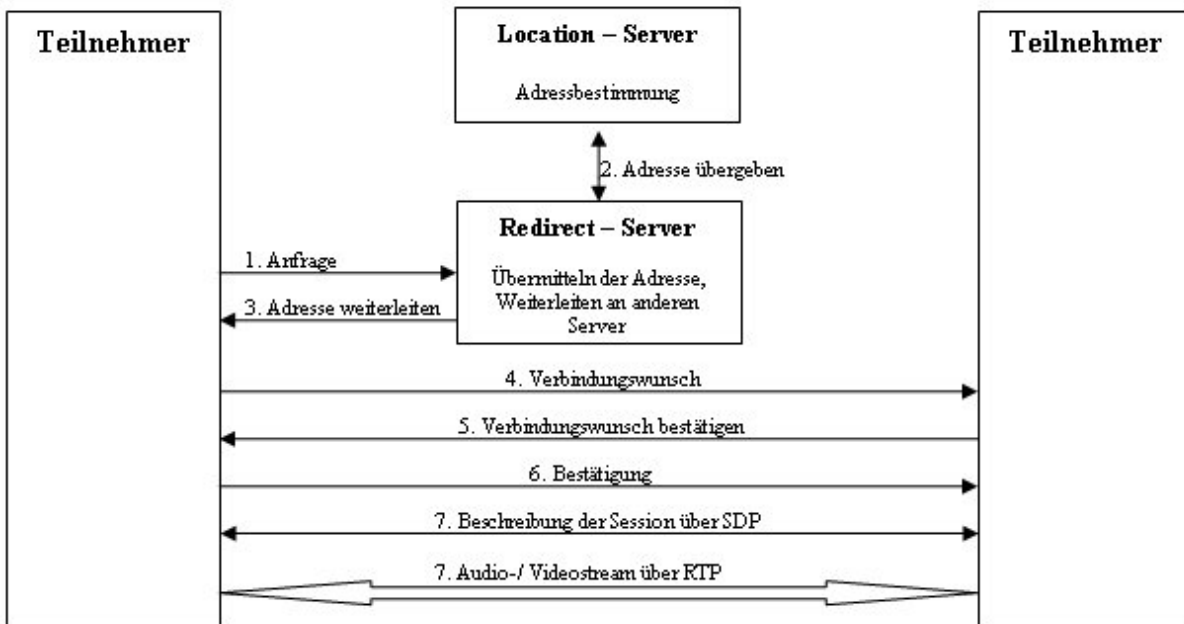


Abbildung 4.16: Verbindungsaufbau über Redirect-Server

4.4 H.323, SIP und Andere im Vergleich

4.4.1 H.323 vs. SIP

Im Gegensatz zu SIP stellt H.323 kein echtes Protokoll dar, es handelt sich um eine Sammlung von vielen Standards und Subprotokollen. Die Struktur von H.323 ist vertikal, das heißt, die festgelegten Protokolle bauen aufeinander auf und sind von einander abhängig. Dadurch wird H.323 unflexibel und schwer erweiterbar. Die SIP-Architektur hingegen ist modular aufgebaut. SIP selbst kommt mit wenigen Nachrichten aus, zur Unterstützung wurden lediglich SDP und SAP entwickelt. SIP kann jedoch mit vielen möglichen Protokollen zusammenarbeiten, ohne dabei von einzelnen abhängig zu sein. Ähnlich sieht es bei den verwendbaren Sprach- und Video-Codecs aus.

Durch die Verwendung von SIP findet keine Einschränkung über mögliche Codecs statt. Hier sind alle durch die IANA Registrierten nutzbar und können durch SDP zwischen den verschiedenen Teilnehmern ausgehandelt werden. Bei H.323 können nur von der ITU standardisierte Codecs genutzt werden.

SIP kann sowohl über TCP als auch über UDP transportiert werden. Durch die Verwendung von UDP können Server stark entlastet werden. Die SIP-Server sind nur zum Auf- und Abbau einer Session nötig, während in einer H.323-Architektur für die gesamte Dauer eines Gesprächs alle Informationsströme über den Gatekeeper laufen. Er muss das Routing von Gesprächen übernehmen und hält dabei eine Vielzahl von TCP-Verbindungen aufrecht, welche das Netz stark belasten können.[7] Der Verbindungsaufbau selbst ist bei SIP ebenfalls viel einfacher und schneller, da eine SIP-Anfrage bereits alle notwendigen Informationen enthält. Bei H.323 ist durch die Verwendung der Protokolle H.225.0, Q.931, H.245 zum Aufbau von Verbindungen eine große Zahl von Round-Trips notwendig, die zu entsprechenden Verzögerungen führen kann.[7]

Nachteilig gegenüber SIP ist bei H.323, dass es nicht multicastfähig ist. Konferenzen müssen immer durch eine MCU gesteuert werden. Sie ist bei diesem zentralistischen Konferenzmodell unverzichtbar, während sie bei SIP durch die Multicastfähigkeit und damit dezentralen Ansatz nicht benötigt wird.

SIP ist textbasiert, während die Signalisierung bei H.323 binär ist. Damit ist SIP leichter zu implementieren, da keine aufwendigen Codegeneratoren nötig sind.[7] Eine SIP-Adresse ist die einer Emailadresse ähnlich. Sie ist nach dem Prinzip user@host aufgebaut. Der User-Teil kann dabei eine Telefonnummer oder ein Name sein, während der Host-Teil ein Domain-Name oder eine Netzwerkadresse darstellt. Zur Auflösung der SIP-Adresse zur IP-Adresse bedient sich SIP dem Domain Name System(DNS).[8] Bei H.323 muss hingegen in einem Gatekeeper eine Datenbank mit Namenszuordnungen angelegt werden.

Da die Gatekeeper ein zentraler Anlaufpunkt bei H.323 sind, kann man hier Methoden zur Abrechnung von Gesprächen aufsetzen. Da ein SIP-System viel verteilter ist, ist das Billing sehr viel schwieriger. Einzige Möglichkeit wäre, dass der Proxy-Server jedes Gespräch bis zum Ende überwacht.

SIP ist drei Jahre jünger als H.323 und deshalb auf dem Markt nur sehr gering vertreten, daher stützen sich viele Hersteller noch auf H.323 ab.

	SIP			H.323
Architektur	Modular	+	-	Vertikal
Transportprotokoll	TCP, UDP	+		TCP (UDP ist optional)
Protokollkodierung	Text	+	-	Binär (ASN.1)
Spezifikation	Nur Signalisierung			Rahmenprotokoll (alle Kommunikationsaspekte)
Komplexität	Gering	+	-	Hoch
Erweiterbarkeit	Leicht	+	-	Schwer
Codec-Unterstützung	Alle IANA-Codecs	+	-	Nur ITU-Codecs
Namenszuordnung	DNS	+	-	Datenbank
Mehrpunktkonferenzen	IP-Multicast	+	-	MCU
Herstellersupport	Steigend	-	+	Favorisiert
Abrechnung	Nur über Proxy-Server	-	+	Über Gatekeeper

Abbildung 4.17: Vergleich SIP-H.323[21][20]

4.4.2 H.323-SIP Interoperabilität

Für das Interworking zwischen H.323-Komponenten und SIP-Benutzern wird ein SIP/H.323-Gateway benötigt. Es ist auf der einen Seite mit dem H.323-Standards kompatibel, auf der anderen Seite verhält es sich wie ein SIP-Server. Will ein H.323-Terminal nun eine Verbindung aufbauen, so wendet es sich an einen Gatekeeper. Nachdem dieser dem Verbindungswunsch zustimmt, leitet er die Anfrage an den SIP/H.323-Gateway weiter. Nach dessen Zustimmung verfährt das Gateway weiter nach dem SIP-Protokoll und signalisiert das Vorhaben einem Proxy-Server, welcher schließlich den Verbindungsaufbau zum Ziel-Client übernimmt. Wird das Gespräch angenommen, so wird dies dem Gateway durch den Proxy-Server signalisiert. Die Bestätigung wird nun über den Gatekeeper an das H.323-Terminal weitergeleitet. Nachdem die Aushandlung der Sitzungskriterien zwischen Terminal und Gateway, beziehungsweise zwischen Gateway und SIP-Client, vollzogen ist, kann der Austausch der Multimediaströme beginnen. Während dieser auch direkt zwischen den Teilnehmern geschehen kann, läuft die Verbindungskontrolle über das SIP/H.323-Gateway ab.

4.4.3 MGCP, Megaco und Skinny

Das Media Gateway Control Protocol, kurz MGCP, beruht auf dem H.323-Standard und wurde von der IETF beschrieben, um eine bessere Interoperabilität zwischen IP- und leitungsvermittelten Netzen zu erreichen. Es setzt an der Konzeption des Gateways an, welches für die Umsetzung der Signalisierung und Mediaströme zwischen den beteiligten Netzen verantwortlich ist. Die Idee bei MGCP ist die Trennung dieser beiden Aufgaben. Dazu gibt es zwei neue Funktionseinheiten. Diese sind das Media-Gateway, welches zum Übersetzen der Mediaströme dient, und der Media-Gateway-Controller, der die Signalisierung umsetzt. Zur Kommunikation zwischen beiden dient dann das MGCP. Der Hintergrund sind Weiterentwicklungen in Sprachfestnetzen, an die auch Voice-over-IP entsprechend angepasst werden muss. Das Protokoll verliert allerdings mit dem Aufkommen



Abbildung 4.18: Verbindungsaufbau H.323-Terminal und SIP-Client

des Megaco/H.248-Standards an Bedeutung.[19] Megaco wurde gemeinschaftlich von der ITU und der IETF entworfen. Im Unterschied zu MGCP erlaubt es Megaco beim Media Gateway nicht nur zum PSTN beziehungsweise ISDN umzusetzen, sondern auch zu weiteren Netzen, wie zum Beispiel dem ATM (Asynchronous Transfer Mode).[8]

Im Laufe der Zeit haben sich viele Firmen eigene Gedanken zur Funktionsweise von Voice-over-IP gemacht. Zum Beispiel hat sich der amerikanische Hersteller Cisco, ausgehend von seiner weltweit führenden Position auf dem Router-Markt, schon früh mit IP-Netzen zur Sprachübertragung beschäftigt. Dabei hat er bereits vor einiger Zeit eine eigene Architektur entwickelt und dabei das Protokoll Skinny entwickelt. Es handelt sich dabei um eine abgespeckte Variante des H.323-Standards und soll laut Hersteller einfacher und funktionaler sein. Allerdings nutzen nur wenige Anbieter das von Cisco offen gelegte, jedoch nicht international standardisierte Verfahren.[22]

Die vielen unterschiedlichen Protokolle, Standards und Architekturen bringen Interoperabilitätsprobleme mit sich. Die zukünftige Aufgabe besteht also darin, sich gemeinsam auf eine Lösung festzulegen, um die Zusammenarbeit der vielen unterschiedlichen Systeme dauerhaft gewährleisten zu können.

4.5 Zusammenfassung und Schlussfolgerungen

Die VoIP-Technologie bietet viele Vorteile für den Nutzer, dennoch ist sie in der Realität nur bedingt einsatzfähig. Zum einen stößt der Privatanutzer schnell an die Grenzen der Internet-Telefonie, wegen der schlechten Gesprächsqualität. Zum anderen erschweren die vielen Standards und proprietären Lösungen die Interoperabilität der verschiedenen Systeme. Damit kann sich der VoIP-Markt nicht so schnell entwickeln wie prognostiziert. Auch mit der neuen Version des Internetprotokolls, IPv6, und eventueller Erhöhung der Kapazitäten des Internets, kann den Anforderungen von Echtzeitanwendungen, wie Voice-

over-IP, zwar ein Stück entgegengekommen werden, doch kann auch so kein vollständiger Quality-of-Service garantiert werden. Damit bleibt in den Spitzenzeiten immer das Risiko der schlechten Gesprächsqualität bestehen und die Nutzer werden sich genau überlegen, ob sie von ihrer hohen Gesprächsqualität, garantiert durch ihr gewohntes Telefonnetz, abgehen werden. Etwas anderes ist es noch, wenn es sich um das eigene IP-Netz handelt. In diesem Fall können alle Komponenten so konfiguriert werden, dass keine Bandbreitenengpässe mehr entstehen. Dass dies besonders für Unternehmen mit bestehenden Datennetzen interessant ist, wurde bereits oben angesprochen. Dennoch nutzen nicht so viele Firmen wie erwartet diese Technik. Dies kann daran liegen, dass man die Vorteile, welche Voice-over-IP mit sich bringt noch nicht genügend heraus gestellt hat. Auf der anderen Seite stehen natürlich die vielen unterschiedlichen Standards in diesem Bereich. Und um sich bei der Einführung von Voice-over-IP nicht in eine Sackgasse zu begeben, sollten Unternehmen darauf achten, standardkonforme Produkte einzusetzen. Nur so wird Austauschbarkeit, Erweiterbarkeit, Interoperabilität, Zukunftssicherheit und Flexibilität sichergestellt. Es bleibt aber noch offen welcher der richtige Standard ist. Entweder die H.323-Architektur, welche an sich schon in vielen unterschiedlichen Versionen existiert, aber weit am Markt vertreten ist, oder das einfachere SIP-Model, welches nach Ansicht vieler Experten die effizientere Lösung ist, jedoch von den Herstellern noch nicht in dem Maße unterstützt wird, wie es notwendig wäre. Damit ist die Situation einigermaßen zerfahren. Viele Anbieter haben trotzdem schon begonnen, ihr ausschließlich H.323 unterstützendes Equipment durch Geräte auszutauschen, die nur SIP unterstützen. Dadurch sind Netzwerke entstanden, die beide Protokolle parallel einsetzen und dadurch teuer und zeitaufwändig in Unterhalt und Pflege sind.[23] Die meisten Experten sind jedoch der Meinung, dass H.323 in den nächsten Jahren weitgehend durch SIP abgelöst wird. Tendenzen dazu lassen sich bereits erkennen. Populärste Anwendung des H.323-Standards ist vielleicht Netmeeting von Microsoft, aber nachdem Microsoft mit Windows XP einen SIP-Client mitliefert und die nächste Generation von Instant-Messaging-Clients aus Redmond auf SIP aufsetzen sollen, wird dieses Protokoll auch auf dem Desktop immer weitere Verbreitung finden.[23] Dass SIP auf dem Vormarsch ist, zeigt auch die Entscheidung, es für die Signalisierung bei UMTS einzusetzen. Dennoch geht Entwicklung zu einem einheitlichem Protokoll nur sehr schleppend voran, da sich viele Hersteller nicht von ihren auf H.323 aufsetzenden Produkten trennen wollen und können. Daher wird es noch einige Zeit dauern, bis sich Voice-over-IP etwas mehr durchsetzen kann. Bis es soweit ist, gilt es einen kosteneffizienten Migrationspfad zu finden. Ziel muss schließlich eine neue Generation von offenen, verteilten VoIP-Systemen sein, die vollständige gegenseitige Konvertierung der Protokolle und nahtlose Zusammenarbeit bieten.[23] Die Realisierbarkeit dessen, wird die Zukunft von VoIP bestimmen. Wird es nicht erreicht, werden VoIP-Systeme immer nur Insellösungen bleiben und die Technologie wird nie in ernsthafte Konkurrenz zum klassischen Telfonnetz, welches sich auch stetig weiterentwickelt, treten können.

Literaturverzeichnis

- [1] Schulzrinne, H.: SIP Web Page; <http://www.cs.columbia.edu/hgs/sip/>
- [2] Handley, M.; Jacobson V.; Perkins C.: SDP: Session Description Protocol; Internet Draft draft-ietf-mmusic-sdp-new-10.txt; <http://www.ietf.org/internet-drafts/draft-ietf-mmusic-sdp-new-10.txt>, September 2002
- [3] Open H.323 Project; <http://www.openh323.org/>, September 2002
- [4] Lange, Christoph: IP-Telefonie im Trend, in: NetworkWorld 10/2002, 31.05.2002
- [5] Schoblick, Robert: Sprache und Daten in IP-Netzen, in NetworkWorld 12/01, 22.06.01
- [6] Detken, Kai-Oliver: Kampf um Normen, in: NetworkWorld 11/2001, 08.06.2001
- [7] Detken, Kai-Oliver: H.323 versus SIP im VoIP-Umfeld, in: NET 3/2002
- [8] Moos, Roland: Voice over IP Standards und Technik, in: Unterrichtsblätter der Telekom Jg.54 7/2001
- [9] Barth, Dr. Matthias: Internet-Telefonie - Heutiger Status und zukünftige Potentiale, Bundesministerium für Bildung und Forschung, 1999
- [10] Eppeler, Klaus: Telefonieren per Internet Protocol: in Funkschau 14-15/2002, 14.10.2002
- [11] E-Online Kommunikationstechnik: Internet-Telefonie; <http://www.e-online.de/sites/kom/0307041.htm>
- [12] E-Online Kommunikationstechnik: Voice-over-IP(VoIP); <http://www.e-online.de/sites/kom/0503131.htm>
- [13] T-Lan: Marktanalysen; <http://www.t-lan.de/>; 2002
- [14] tecChannel: So funktionieren TCP/IP und IPv6; <http://www.tecchannel.de/internet/209/24.html>
- [15] Universität Koblenz-Landau, Fachbereich Informatik, Institut für Wirtschaftsinformatik: Fax over IP; <http://cti.iwi.uni-koblenz.de/iwi/>
- [16] Universität Koblenz-Landau, Fachbereich Informatik, Institut für Wirtschaftsinformatik: Motivation und Überblick; <http://cti.iwi.uni-koblenz.de/iwi/>

- [17] IMTC; <http://www.imtc.org/>
- [18] Siemens; H.323: <http://w3.siemens.de>
- [19] Siemens; MGCP: <http://w3.siemens.de>
- [20] Zhang, Lanlan; Xu, Xiaochun: H.323 und SIP, Proseminar - Internetprotokolle für Multimediakommunikation, RWTH Aachen, Lehrstuhl für Informatik IV, Wintersemester 2002/2003
- [21] Wolf, Dr. Ludwig; Heik, Andreas: Lehrmaterial - Multimedia Netz Praxis, Technische Universität Chemnitz, 03.02.2003
- [22] Rossenhövel, Carsten; Lange, Christoph: Einzeltest VoIP-Anlage "MCS 7825-800" von Cisco Systems; <http://vd-consult.net/VOIP-Cisco-einzeltest.htm>
- [23] Vassalo, Michael: Vom Konflikt zur Koexistenz, VoIP-Trends H.323 und SIP; <http://www.lanline.de>

Kapitel 5

Content Charging in the Internet

Danny Heerlein

In den Jahren 1997 bis 2000 kam es zu einem sprunghaften Anstieg der Internetnutzer. Um eine schnellere Verbreitung des Internets herbeizuführen wurden Anreize in Form von kostenlosen Serviceleistungen und Inhalten geschaffen.

Ende 1999 und im Jahr 2000 kam es zu einem Zusammenbruch dieses Systems. Immer mehr kostenlose Dienste wurden eingestellt bzw. durch kostenpflichtige Dienste ersetzt. Dieser Wandel verlangt nach neuen Zahlungsmethoden die dem Zahlen im Internet besser angepasst sind.

Eine Analyse der Gründe für diesen Wandel, die unterschiedlichen Arten der Klassifikation vorhandener Zahlungsmodelle, konkrete System und deren Anbieter sind Gegenstand dieser Arbeit.

Inhaltsverzeichnis

5.1	Einleitung	105
5.1.1	Geschichtliche Entwicklung	105
5.1.2	Tarifmodelle im Internet	106
5.1.3	Kostenlos angebotene Serviceleistungen	108
5.2	Vom kostenfreiem zum kostenpflichtigem Service	109
5.2.1	Ökonomische Ursachen	110
5.2.2	Technische Schwächen kostenloser Services	111
5.2.3	Beispiele für kostenpflichtige Services	114
5.3	Zahlungsmodelle	116
5.3.1	Anforderungen	117
5.3.2	Klassifikation Digitaler Zahlungssysteme	119
5.3.3	Zahlungsmodelle	126
5.3.4	Existierende Systeme	128
5.3.5	Anbieter von Zahlungssystemen	132
5.3.6	Analyse potenzieller Modelle auf Akzeptanz	133
5.4	Zusammenfassung und Ausblick	134

5.1 Einleitung

In Deutschland ist es momentan üblich seinen Internetzugang durch einen Service Provider zur Verfügung gestellt zu bekommen. Die meisten Service Provider bieten hierbei verschiedenen Tarifmodelle, die nachfolgend erklärt werden, an. Die Nutzung verschiedener Serviceleistungen im Internet, wie zum Beispiel:

- Suchmaschinen
- Newsarchive
- Downloadbereiche für Musik- bzw. Videodaten

ist weitgehend kostenlos.

Allerdings lässt sich ein deutlicher Trend zum Wandel dieser kostenloser Service hin zu kostenpflichtigen Services erkennen. Verschiedene Serviceanbieter haben diesen Wandel bereits vollzogen oder sind im Übergang.

Es scheint so, als ob der Endkunde bereit ist für in Anspruch genommene Serviceleistungen im Internet zu bezahlen.

Die Ursachen für diesen Wandel und die damit verbundene Frage nach Zahlungsmodellen im Internet und deren Akzeptanz bei dem Endkunden sind Gegenstand dieser Arbeit.

5.1.1 Geschichtliche Entwicklung

Das Internet ist eine Netzwerkstruktur die auf der Grundlage der TCP/IP Protokollfamilie, Anfang der 80iger Jahre in den USA durch die Verbindung von Universitäts-, Regierungs- und Militärnetzen entstanden ist. Sie basiert auf Absprachen und Einzelverträgen über die Nutzung von bestimmten Netzteilen. Das Internet ist somit keine Gesellschaft oder Firma, sondern ein auf Verträgen und Übereinkünften technischer und organisatorischer Art basierender Netzzusammenschluß. Zur Koordinierung der Adreßräume sind lediglich die NICs (Network Information Center), z. B. das InterNIC für die gesamte Erde, das RIPE für den europäischen Raum und das DE-NIC für Deutschland zuständig, die gegen vergleichsweise geringe Gebühr Netzwerk-Adressen offiziell registrieren. Standards zur Kommunikation der Netzteilnehmer werden im Netz diskutiert und in RFCs veröffentlicht. Das Netz lebt somit auch von seinem wissenschaftlichen und universitären Umfeld, aus dem viele Weiterentwicklungen unentgeltlich kommen. Ein Anschluß an das Internet der ersten Jahre war nur durch teure exklusive Standleitungen möglich. Deshalb beschränkte sich der Nutzerkreis auf Universitäten und große Firmen. Wenige große Nutzer schlossen Verträge, deren Tarife nur von der in Anspruch genommenen Bandbreite abhing.

Die Keimzelle des kommerziellen Internets in Deutschland ist die Universität Dortmund, die als erste seit 1985 über das EUNet Internet an kleinere kommerzielle und private Kunden vermarktet. In den folgenden Jahren kamen einige Anbieter hinzu, die nach dem gleichen Prinzip den Zugang ermöglichen: Die Anbieter, im folgenden auch Provider genannt,

mieten Leitungen zu den großen Backbones (Netzverbindungen der Provider mit großer Bandbreite) in den USA oder Europa und bündeln diese Anschlüsse an einem Standort in Deutschland. Damit haben die Provider meist reine Fixkosten für Ihre Anbindung zu zahlen. Von dort aus werden Wähl- oder meist Standleitungen zu Vertragshändlern, im folgenden PoPs (Points of Presence) genannt, geschaltet. Diese betreuen die Kunden direkt vor Ort und können Zugänge zum Ortstarif anbieten.[1]

5.1.2 Tarifmodelle im Internet

Es haben sich im Laufe der Zeit eine ganze Reihe von Tarifarten gebildet, die sich an den Bedürfnissen der neuen Nutzer und an den technischen Gegebenheiten orientieren. Vergleiche von Preisen und Leistungen zwischen den verschiedenen Anbietern gestalten sich für den Nutzer äußerst schwierig. Meist haben die Kunden bei der Suche nach einem Internetzugang auch keine genaue Vorstellung, was sie erwartet und wie sie das Netz nutzen werden. Die Vor- und Nachteile der unterschiedlichen Tarifmodelle werden im Folgenden dargelegt.

Pauschale Tarifierung

In Anlehnung an die ursprünglich vorherrschenden Mietleitungen bieten viele ISP einen Internetzugang zu einem pauschalen monatlichen Festbetrag an (flat rate). Falls der PoP auch mit seinem Provider pauschal abrechnet, ist dieser Tarif für den PoP und für seine Kunden gleichermaßen kalkulierbar und damit risikoarm.

Der PoP hat fixe monatliche Kosten durch Personal, Miete und Abschreibung, die durch die ebenfalls fixen Einnahmen gedeckt werden. Ein Aufwand für die Messung der tatsächlich in Anspruch genommenen Leistungen durch Nutzung des Internets entsteht nicht, da eine solche Messung nicht stattfindet. Die Kosten für die Rechnungserstellung sind bei dieser Art von Tarifierung auf ein Minimum beschränkt.

Viele Kunden wünschen sich einen pauschalen Tarif oder bestehen sogar auf einem pauschalen Tarif, da sie andernfalls die ihnen entstehenden Kosten nicht einschätzen können. Insbesondere Informationsanbieter (WWW und ftp) vermuten zu Recht ein großes Risiko, da durch eine unerwartet viele Zugriffe auf ihre Dateien hohe Kosten für sie entstehen können. Ein Pauschalpreis schützt sie vor unangenehmen Überraschungen.

Der Nachteil dieser Abrechnungsart ist (aus der Sicht des Kunden betrachtet), daß zwar die Kosten für ihn kalkulierbar werden, nicht jedoch die Qualität der Leistung, die sich in der Erreichbarkeit von Adressen und in der Höhe des Datendurchsatzes ausdrückt. Das Interesse des PoPs bei pauschalen Tarifen ist es, möglichst viele Kunden zu gewinnen, die sich dann die begrenzte Bandbreite teilen müssen. Bei den Kunden gibt es keinen Grund, mit der Verwendung des Mediums sparsam umzugehen. In der Praxis geht deshalb ein pauschal tarifizierter Zugang meist mit schlechter Qualität einher. Manche PoPs werben mit ihrer Bandbreite zum Kunden und machen sogar die Höhe ihres pauschalen Tarifs von dieser Bandbreite abhängig. Das ist deshalb unseriös, weil damit dem Kunden überhaupt nicht die viel entscheidendere Frage klar wird, mit wieviel anderen Kunden er sich die Standleitung zum Provider teilen muß, durch die die Geschwindigkeit seines Datenflusses determiniert ist.[1]

Zeitabhängige Tarifierung

Eine Abrechnung nach Zugangszeit ist für den PoP technisch einfach zu realisieren und für den Kunden auf den ersten Blick transparent, da er die Nutzungszeit selbst bestimmt und bei Nichtnutzung auch nichts bezahlt.

Je schlechter der angebotene Datendurchsatz für den Kunden ist, desto mehr muß er bezahlen, da für die Übermittlung einer bestimmten Datenmenge mehr Zugangszeit in Anspruch genommen wird. Der Kunde hat nicht nur den Kostennachteil, der sich aus überlasteten Leitungen ergibt, sondern auch lästige Wartezeit am Rechner. Aus diesen Gründen kommt die zeitabhängige Abrechnung nur für den gelegentlichen Nutzer in Frage. [1]

Volumenabhängige Tarifierung

Der älteste und in Deutschland immer noch am meisten verbreitete Tarif zwischen Provider und PoP ist der volumenabhängige Tarif. Hierbei wird gesendetes wie empfangenes IP-Volumen gleichermaßen gemessen und in Rechnung gestellt. Lokaler Verkehr zum Beispiel:

- lokalen ftp-Servern
- zum Domain-Name-Service
- zu News-Servern

wird meist schon aus technischen Gründen nicht erfaßt und tarifiert. Einige Provider unterscheiden bei der Tarifierung nach nationalem und internationalem Verkehr.

Wenn der Provider nur diese Art der Abrechnung mit den PoPs anbietet, sind auch die PoPs auf volumenabhängige Tarifierung festgelegt, da sie das Risiko eines pauschalen Angebotes an ihre Kunden scheuen. Der Erfassungsaufwand ist bei dieser Tarifierung am größten und muß im Endeffekt auch vom Kunden mitbezahlt werden.

Für Kunden, die Informationen im Netz anbieten wollen, sind die Kosten schwer kalkulierbar. Eine etwaige Trennung zwischen nationalem und internationalem Verkehr verschärft das Problem. Kunden haben auch die berechtigte Sorge, daß ihnen sehr große, möglicherweise sogar nutzlose E-Mails zugeschickt werden, deren Volumen sie dann bezahlen müssen. Aus technischer Sicht ist dies nicht vermeidbar.

Merkmal dieser Abrechnungsart ist, daß sie ein Kostenbewußtsein im Hinblick auf die Bandbreite schafft, welches sich günstig auf die Qualität auswirkt. Der PoP hat ein Interesse daran, eine möglichst große Bandbreite anzubieten, um an dem größeren Datenvolumen zu verdienen. Der Kunde bekommt auf diesem Wege gute Reaktionszeiten, was gerade bei Diensten wie WWW unabdingbar notwendig ist. Der Kunde bezahlt oft bei dieser Tarifierung etwas mehr, erhält jedoch dafür eine bessere Qualität.

Die volumenabhängige Abrechnung erfordert ein gewissenhaftes Erfassen des Datendurchsatzes, welches jedoch durch die technischen Randbedingungen nicht immer leicht zu gewährleisten ist. Abhängig von der eingesetzten Hard- und Software kommt man leicht zu unterschiedlichen Abrechnungsergebnissen. [1]

Mischformen

Auf dem Markt gibt es diverse Mischformen, die die oben genannten Abrechnungsarten in verschiedener Weise kombinieren.

Oft werden pauschale monatliche Grundgebühren erhoben, in denen ein Freikontingent an Volumen bzw. Zeit enthalten ist. Alles was darüber hinaus anfällt, wird volumen- bzw. zeitabhängig tarifiert. Viele PoPs tarifieren zusätzlich zum Volumen auch noch die Zugangszeit, um ebenfalls hier ein Kostenbewußtsein beim Kunden zu schaffen.

Pauschal anbietende PoPs beschränken häufig Privatpersonen, bei entsprechend günstigeren Tarifen, die Zugangszeit auf die Abend- und Nachtstunden, um am Tage Firmen und anderen „Vollzahlern“ einen besseren Datendurchsatz zu bieten. Es gibt auch das Verfahren, die Gesamtstundenzahl des Zugangs pro Monat einzuschränken.

Einige PoPs haben im Hinblick auf das Volumen gestaffelte Größenklassen gebildet, für die entsprechend gestaffelte Pauschalpreise gelten. Solange der Kunde sich innerhalb einer bestimmten Größenklasse bewegt, hat er den Vorteil eines festen und damit für ihn kalkulierbaren Betrages.

Pauschal werden in der Regel alle fest definierbaren Dienstleistungen abgerechnet wie z.B. die Einrichtung von Netzen. [1]

5.1.3 Kostenlos angebotene Serviceleistungen

In diesem Abschnitt soll dem Leser verdeutlicht werden, wie es zu einem solch imensen Angebot an kostenlos angebotenen Diensten im Internet kam und warum sich diese Dienste größtenteils nicht etablieren konnten.

Dot Com Boom

In den Goldrauschjahren des Internets, zwischen 1997 und 2000 war Portabilität bei den bekanntesten Internet-Firmen (auch Dot Coms genannt, da sie zum größten Teil eine .com Domain hatten) eine Nebensache.

Die Strategie der meisten dieser Dot Coms war auf nur eines gezielt: einen rasanten Wachstum durchzuführen, um dann den gezielten Markt bzw. die gezielte Marktnische zu beherrschen. Dieses hohe Wachstumstempo erforderte Investoren mit viel Kapital und unwiderstehliche Anreize für die Kunden (oder Zielgruppe).

Viele sahen im Internet einen Wirtschaftsmotor. Neue Begriffe wie E-Commerce und New Economy waren in Venture Capital floß in Strömen. Die Hoffnungen und Erwartungen waren hoch. Niemand wollte den Internet-Zug verpassen. Die Aktien der Dot Coms waren alle überbewertet, doch das störte Niemanden, denn das Internet war die Zukunft.

Neben der enormen Kapitalspritze war Anziehungskraft die zweite Säule der Strategie des rasanten Wachstums und diese wurde hauptsächlich mit kostenlosen Diensten und Produkten erreicht.

Alles war kostenlos von E-Mailadressen, Faxnummern, SMS, Chat, Diskussionsforen, Webhosting, Software, Verzeichniseinträge... Kostenlose Dienste waren unwiderstehlich, und die Besucher kamen in Strömen.

Niemand wußte so genau, wie man aus dem enormen Besucheransturm richtiges Geld verdienen konnte, aber das war den meisten auch egal, denn sie bauten ihre gesamte Zukunft auf Werbeeinnahmen - hauptsächlich auf Bannerwerbung.[2]

Dot Com Crash

Ende 1999 und durch 2000 begann der Dot Com-Boom zu schwächen, um dann ganz zusammenzubrechen. Firmen gingen täglich pleite oder wurden von profitablen Konkurrenten aufgekauft. Wie kam es zu diesem jämmerlichen Ende?

Mehrere Faktoren waren da im Spiel. Investoren wollten endlich Profite sehen, doch viele der Dot Coms konnte keine liefern.

Die vielversprechende Bannerwerbung war auf eine Spirale nach unten. Banner CTR-Zahlen (CTR = Click-Through Rates = die Anzahl der Leute, die auf den Banner klickt um auf der Website des Werbers zu gelangen) waren im Keller. Werbeeinnahmen sanken auf ein Niveau runter, so daß viele Websites, gerade Content-Sites (Websites, die ausschließlich kostenlose Inhalte boten und sich ausschließlich von Bannerwerbungen ernährten) pleite gingen.

Der Fall der Dot Coms löste eine Lawine in den Aktienmärkten aus. Technologie-Werte, die zum größten Teil aus Internet- oder Biotech-Firmen bestehen, sanken von einem Rekord Tief zum anderen.

Das Internet und die Dot Coms waren nicht mehr Symbole einer neuen Zukunft - eher Reste der blinden Gier.

Und jetzt?

Jetzt wirken die Internet-Firmen wie normale Firmen. Sie bauen auf realistische (anstatt phantastische) Business-Pläne und konzentrieren sich auf die Erzielung von Profiten, anstatt rasanten Wachstums. Selbst der Begriff Dot Com ist out, so out sogar, daß sich viele Firmen das Dot Com aus Ihren Namen streichen ließen.

Die Website theendoffree.com zeigt einen aktuellen Stand der Dienste, die früher kostenlos waren und jetzt kostenpflichtig wurden.

Das Internet ist nicht mehr New Economy, sondern einfach Economy, wie alle anderen Wirtschaftszweige. [2]

5.2 Vom kostenfreiem zum kostenpflichtigem Service

Im vorangegangenen Kapitel wurde dargestellt wie es zu einer Entwicklung von kostenlosen Diensten im Internet kam. Fehlende Werbeeinnahmen wurden als Hauptgrundgrund für einen Wegfall dieser genannt. Parallel zu diesen ökonomischen Ursachen muss man jedoch auch die technischen Schwächen der kostenfreien Services betrachten.

5.2.1 Ökonomische Ursachen

Immer mehr Anbieter erkennen nach Jahren des Booms, dass kostenlose Angebote im Web nicht finanzierbar sind. Reine Werbe-finanzierte Dienste verschwinden nach und nach von der Bildfläche. Beliebte Web-Seiten gehen zunehmend dazu über, für ihre Leistungen von jedem Nutzer Geld zu verlangen. Doch die digitalen Zahlungsmöglichkeiten sind noch nicht voll ausgereift.

Lange war es bekannt: auf Banner wird nur noch in 2% aller Fälle geklickt. Und doch wollte eine ganze Branche ausgerechnet damit ihr Geld verdienen. Das konnte doch nur schiefgehen. Der Fall kam früher als gedacht, die Branche steckt in der Krise.

FAZ, T-Online, Britannica.com und Web.de sind einige von den großen, die mit dem Gedanken spielen, sich ihre Inhalte und Dienste künftig von den Nutzern bezahlen zu lassen. Napster ist ein anderes populäres Beispiel, wo sich die Freiheit im Internet nicht mit wirtschaftlichen Überlegungen vereinen lässt. Unter den Anbietern kostenlosen Speichers, die bei Erscheinen von StarHTML 3 im März 2000 gelistet waren, haben mindestens 5 inzwischen den Dienst aus Finanzierungsgründen eingestellt.

Service-Anbieter die sich nicht durch Werbeeinnahmen finanzieren mussten lebten von Förderungen. Zum Beispiel wurde das Internet-Angebot der „Hamburger Morgenwelt“ durch eine Förderung der Heidelberger Klaus-Tschira-Stiftung finanziert.

T-Online-Chef Holtrop am 20.04.01: „Nutzer müssen künftig für bestimmte Inhalte zahlen“

Ich bin der Meinung, dass sich eine inhaltsbezogene Abrechnung im Internet zwar durchsetzen, aber kostenlose Dienste nicht verdrängen wird. Eine einfache Möglichkeit der Abrechnung und erweiterte Dienste werden diesen Wandel zwar beschleunigen aber nicht vervollständigen. So lange ein Teil der Endnutzer die Schwächen kostenloser Dienste in Kauf nehmen und diese Dienste nachfragen werden diese auch angeboten.

Nach Flatrate-Fiasko und Börsen-Debakel bei T-Online glaubt Thomas Holtrop, der Vorstandsvorsitzende von Deutschlands größtem Online-Provider, nun das Erfolgsrezept für sein Unternehmen in der Hand zu haben: „Wir müssen von der Kostenlos-Kultur im Internet wegkommen.“ Gegenüber der Zeitschrift ONLINE TODAY äußert er sich erstmals in einem längeren Interview: „Wenn wir z.B. ein Konzert mit Eric Clapton präsentieren, wird es bald so sein, dass ab einer bestimmten Minute die Gebühreuhr tickt. Die bisherige Praxis, dass Inhalte umsonst angeboten werden, ist ökonomisch nicht durchzuhalten. Der User muss bereit sein, einen Preis für Mehrwert zu zahlen.“ [5]

Der Grund für das gescheiterte Flatrate-Projekt ist für Holtrop eindeutig: „Manche User waren täglich 20 Stunden online, sie nahmen die Flatrate quasi als Standleitung. Dies war technisch und wirtschaftlich nicht tragbar.“ T-Online müsse für die Flat-Kunden minutenbasierte Tarife an den Mutterkonzern Deutsche Telekom entrichten. „Das kann bei diesem hohen Minutenaufkommen die Eigenkosten nicht decken“, so Holtrop. (Vergleich Tarifmodelle Kapitel 1)

Allerdings ist T-Online nicht das einzige Unternehmen, das mit diesen Problemen zu kämpfen hat. Alle Flatrate-Anbieter können ein Lied davon singen

- und für viele von ihnen war es im vergangenen Herbst das Lied vom Tod. Am epidemieartigen Flatrate-Sterben war der Mutterkonzern von T-Online, die Deutsche Telekom auch nicht ganz unschuldig, schließlich hat sie sich bisher nicht bereit gefunden, ihren Mitbewerbern ein akzeptables Großhandelsangebot für Internetzugänge zu unterbreiten.[4]

5.2.2 Technische Schwächen kostenloser Services

1. **Sicherheit** Anbieter für Serviceleistungen verlangen für den Zugang zum System meist persönliche Informationen vom Besucher. Häufig muß der Surfer erst einen Fragebogen ausfüllen, bevor er einen Dienst nutzen kann. Die Betreiber wollen so ziemlich alles wissen: Name, EMail-Adresse, Geschlecht und Alter, häufig sogar Hobbys, Offline-Adresse, Ausbildungsstand und Gehalt.

Einige Betreiber mißbrauchen zumindest die EMail-Adresse der Kunden für die elektronische Form von Postwurfsendungen, Spam genannt. Verfügen sie einmal über die Adresse, so findet der Benutzer regelmäßig Informationen über „Produktverbesserungen“ oder irgendwelche andere Werbung in seinem Postfach, die in überhaupt keinem Zusammenhang mit dem Dienst steht. Mit EMail-Adressen wird ein reger Handel getrieben.

In fremde Hände:

Einige kostenlose Dienste bergen massive Risiken in Bezug auf den Datenschutz. Ein Surfer, der seine Adressen online, etwa mit Yahoo Address verwaltet, gibt vertrauliche Daten Dritter in fremde Hände. Wer würde schon sein Adreßbuch einem fremden Passanten in die Hand drücken? Noch bedenklicher sind Backup-Dienste wie @Backup oder der Online-Virenschanner von Housecall - immerhin hat hier irgendein Fremder Zugriff auf die gesamte Festplatte des PC!

Selbst Dienste, die auf den ersten Blick harmlos erscheinen, etwa Online-Bookmark-Verwalter wie Oneview, bergen ein Risiko: Ein direkter Verweis zum Online-Konto verirrt sich beispielsweise schnell mal in die Linksammlung. Abgesehen davon liefert ein URL-Katalog ein ziemlich exaktes Persönlichkeitsprofil des Surfers. Nicht nur Werbetreibende lecken sich nach solchen Informationen alle Finger.

Internetseiten die mit brisanten Daten hantieren, also beispielsweise Backup-Dienste oder WWW-Adreßbücher, dürften ein beliebtes Ziel von Hackern sein. Dieses Problem tritt natürlich nicht nur bei kostenlosen Anbietern auf. Aber aufgrund der fehlenden Gebühren für diesen Service stehen dem Anbieter weniger Möglichkeiten zur Verfügung diesen sicher zu gestalten. Also ist die Wahrscheinlichkeit einer Sicherheitslücke bei einem kostenlosen Service im Mittel höher als bei einem Kostpflichtigen.

Hier zeigt sich ein generelles Problem im Internet. Im Unterschied beispielsweise zu einer Bank, der man sein Geld anvertraut, haben sich noch keine Mechanismen herausgebildet, die einem Anwender Vertrauenswürdigkeit und Sicherheit garantieren. Nach einem Banküberfall ist das Geld der Kunden nicht weg; ein Einbruch oder Mißbrauch beim Portal-Betreiber wirken sich aber direkt auf die Daten der Anwender aus. Zudem ist nicht einmal gesagt, daß man es bemerkt. Denn den Daten sieht man es schließlich nicht an, wer sie gelesen hat.[3]

2. **Verfügbarkeit** Hier geht es besonders um die Lebensdauer der Internet-Dienste. Es steht zu befürchten, daß einige der Dienste auf Dauer nicht überleben werden. Viele Betreiber kostenloser Online-Angebote machen derzeit noch ein Minusgeschäft und hoffen, irgendwann einmal schwarze Zahlen zu schreiben. Nach einer Untersuchung der Universität Eichstätt ist es bisher keiner deutschen Tageszeitung gelungen, mit ihrem WWW-Angebot einen Gewinn zu erwirtschaften. Sie hoffen vielmehr, durch ihren Einstieg ins Online-Geschäft möglichst früh in einem zukunftssträchtigen Markt dabeizusein. Man kann sich also nicht darauf verlassen, daß eine bestimmte Funktion im Internet auf immer verfügbar ist. Wenn ein vielbenutzter Dienst die Pforten schließt, entspricht das in etwa einem Festplattencrash auf dem lokalen PC. Wer also etwa mit einem der Online-Services seine Adressen pflegt, sollte - wie beim Desktop-PC auch - immer ein Backup seiner Adreßbestände pflegen.[3]

3. **Auslastung** In einer Studienarbeit der technische Hochschule Aachen von von 1994 wird bereits der Aspekt der zunehmenden Auslastung kostenlos angebotener Serviceleistungen als ein möglicher Grund für eine Kommerzialisierung selbiger aufgeführt.: „Es stellt sich die Frage, welchen Einfluß die zunehmende Zahl von Nutzern auf Inhalt und Kultur des Internets haben wird. Unter dem Aspekt des globalen Wissensspeichers wird eine große Zahl von Nutzern vorteilhaft sein. Je mehr Experten zu den verschiedensten Themen am Netz teilnehmen, um so eher wird man eine Antwort auf seine Fragen finden. Für Mailing-Listen und NetNews ist es dabei unerheblich, wie viele diese Dienste nutzen.

Anders sieht es aus bei Informationsservern, die oft in freiwilliger Arbeit betrieben werden und bezüglich der Zugangskapazität ihre Grenzen haben. Eine Explosion der Nutzerzahlen, wie sie derzeit stattfindet, wird zu einer Überlastung einiger Server und zur Einstellung oder Kommerzialisierung bisher freiwilliger Dienste führen.

Schon heute sind viele Dienste durch die Vielzahl der Nutzer überlastet, die wenigen verbliebenen Nischen außerhalb der Bürozeiten werden bald auch von privaten Anwendern ausgefüllt werden. Die Entwicklung hin zu Multimedia-Anwendungen wird ein übriges tun, die Netze bei den heutigen Übertragungskapazitäten vor Probleme zu stellen. Die ständig anwachsende Datenmenge stellt auch die Suchmaschinen vor Probleme. Bei einem Umfang von jetzt 1,3 GB nur für die Titel der Gopher-Menüs sehen sich schon die Veronica-Betreiber dazu veranlaßt, Richtlinien zur Indizierung von Servern zu empfehlen, um die Datenmenge in Grenzen zu halten.

Einziger Ausweg ist die Verbesserung der Infrastruktur, die aber so hohe Kosten verursacht, daß sie nicht mehr wie bisher ausschließlich durch staatliche Mittel finanziert werden kann. Die notwendige Einbeziehung kommerzieller Geldgeber wird sicherlich einen nachhaltigen Einfluß haben.“ [?]

Im Endeffekt stellt sich sich folgendes Problem dar:

Wenn eine gute Serviceleistung angeboten wird, so wird sie auch auf zunehmende Akzeptanz bei verschiedenen Nutzergruppen stossen. Folglich wird auch die Anzahl derer die diesen Service nutzen sprunghaft ansteigen, was wiederum zu einer höheren Auslastung der eingesetzten Technik führt. Dem Anbieter bleiben im wesentlichen nur zwei Möglichkeiten:

- eine Erweiterung der Technik Schaffung von Redundanz bis hin zu Lastenverbänden einer grösseren Zahl von Nutzern gleichzeitig diese Leistung anbieten zu können.

- eine Begrenzung der Zugänge Also die Zulassung einer maximalen Nutzerzahl. Ist diese erreicht müssen sich neue Nutzer solange in einer Art Warteschlange gedulden, bis sich andere Nutzer auswählen.

Gehen wir zunächst auf die zweite Variante ein.

In der heutigen Zeit, wo es praktisch kein Monopol mehr auf einem Gebiet der Serviceleistung gibt, wird ein Kunde der sich subjektiv gesehen zu lange in der Warteschlange aufhält lieber einen anderen vergleichbaren Dienst nutzen. Langfristig gesehen wird der Anbieter dieser Serviceleistung also Kunden verlieren. Was wiederum dazu führt, dass er Werbeaufträge nicht mehr einhalten kann und diesen Dienst nicht mehr komplett kostenfrei zur Verfügung stellen kann.

Nun zur ersten und von den meisten Anbietern auch gewählten Variante.

Die Anschaffung zusätzlicher Technik entspricht zwar einem einmaligen und somit auch kalkulierbaren finanziellen Aufwand allerdings verursachen die damit verbundenen Wartungen, die Kosten für zusätzlich benötigte Bandbreite etc. weitere permanent anfallende Kosten die sich dann meistens nicht mehr allein durch Werbung finanzieren lassen.

Am Beispiel von „oneShell Internet Services GmbH“ möchte ich die Vorteile kostenpflichtiger Dienste im Internet einmal kurz darstellen.[12]

Die „oneShell Internet Services GmbH“ besteht seit Frühjahr 2001. Anfänglich hauptsächlich auf die Bedürfnisse von Geschäftskunden im Rahmen großer Websites und E-Commerce Applikationen spezialisiert, wurden nach kurzer Zeit aufgrund der hohen Nachfrage auch Angebote für Privatkunden in das Kerngeschäft aufgenommen.

Die Server dieses Serviceanbieters sind in einem der besten Rechenzentren Europas der Firma „Global-Switch“ in Frankfurt, mit einer mehrfach redundanten Glasfaserverbindung an das Internet angebunden. Somit kann dieser Dienstleister für jede Anwendung und jeden Kunden ausreichend skalierbare Bandbreite garantieren, die benötigten Kapazitäten sind jederzeit mehrfach vorhanden und stoßen dank des ständigen Netzwerkausbaus nie an ihre Grenzen. Durch dreifach redundante Notstromaggregate mit Dieselmotoren und Großbatterien die eine unterbrechungsfreie Stromversorgung garantieren sowie die Argon-Gaslöschanlage welche jeden Brand innerhalb von Sekunden rückstandsfrei löscht ist eine maximale Ausfallsicherheit garantiert.(siehe Schaffung von Redundanz im vorangegangenen fiktiven Beispiel)

Durch Direktverbindungen zu den größten Internet Providern (z.B. auch der Telekom) kann der Anbieter dieser Serviceleister Pings von maximal 30ms bis zum jeweiligen Provider garantieren. Abhängig vom jeweiligen lokalen Provider liegen die Ping-Werte dann bei ca. 10ms bis 80ms, je nach Art der vom Benutzer verwendeten Internetverbindung. Die Netzwerkverbindungen werden von einem ständig einsatzbereitem Network Operations Team überwacht und bei Problemen automatisch auf alternative Netzwerkverbindungen umgeschaltet. Hierbei wird ständig kontrolliert dass automatisch die schnellste Internetverbindung gewählt wird. Für Datenbanken und Webauftritte werden schnelle MySQL-Datenbanken die mehrmals täglich durch Backups gesichert werden sowie speziell optimierte Apache-Webserver verwendet. Die Vorteile dieses kostenpflichtigen Services noch einmal Stelvertretend für alle anderen kostenpflichtigen Serviceleistungen zusammengefasst:

- Schaffung von Lastenverbänden
- Keine Auslastung der technischen Möglichkeiten um sich so Reserven zurückzuhalten
- Ausfallsicherheit durch Redundanz
- Garantie von Dienstgüteparametern
- Unterhaltung eines Service-/ Wartungsteams
- Durch regelmäßige Backups ist der Großteil der Daten vor Verlust geschützt

Die jeweiligen Preise interessieren hier nur zweitrangig und können auf der HP eingesehen werden.

5.2.3 Beispiele für kostenpflichtige Services

Newsarchiv der NNZ

Während die Aktuelle Ausgabe (auch die Sonntagsausgabe) im Volltext kostenlos zur Verfügung gestellt wird, ist die Benutzung des Newsarchives des Recherchendienstes kostenpflichtig.

Der Preis pro Artikel beträgt 2.00 Euro, der minimale Rechnungsbetrag ist auf 4.00 Euro angesetzt. Bezahlt werden kann bei einmaliger Benutzung direkt mit Kreditkarte (American Express, Visa, Eurocard) oder als registrierter GBI Profikunde durch monatliche Rechnungsstellung.[13]

GBI the contentmachine

[14] „GBI the contentmachine“ ist Datenbankproduzent, Datenbankhost und Informationsdienstleister unter einem Dach. Durch Integration von Content in die GBI Datenbankangebot werden für Verlage weit reichende Synergien bei der Marktbearbeitung erschlossen. Kunden sind u.a. Großunternehmen, Beratungsfirmen, Banken, Versicherungen und Verlage.

GBI setzt 10.000 professionelle Rechercheuren ein. Ein monatlicher Zugriff von über 2,5 Millionen Besuchern der Website lässt sich verzeichnen.

„GBI the contentmachine ist auf dem deutschen Datenbankmarkt - der lange als unprofitabel galt - seit über 20 Jahren erfolgreich tätig. Im Jahr 2001 haben wir mit Datenbankproduktion und Hostbetrieb 6 Mio. EUR Umsatz erzielt. Unser Erfolgsrezept: Kompetenz, Phantasie und ein Gespür für die Bedürfnisse von Kunden und Nutzern.“ GBI arbeitet gewinnorientiert und ist unabhängig von staatlicher Förderung oder dem Einfluss großer Konzerne.

Das Informationsangebot in Zahlen:

- 300 Datenbanken
- 12 Millionen Unternehmensdaten
- 20 Tages- und Wochenzeitungen
- 6 Millionen Literaturhinweise
- 200 Fachzeitschriften
- 50.000 Marktstudien
- 1,5 Terabyte Daten
- 20.000 Sprüche und Aphorismen
- 5 Millionen Infos über Personen
- 9.000 ausgewertete Zeitschriften

Als GBI Profikunde registrierte Nutzer erhalten für eine Servicegebühr von 13.00 Euro pro Kalenderjahr folgende Zusatzleistungen:

- persönliches Login/Username*
- monatliche Rechnungsstellung (Belastung auf Kreditkartenkonto, Lastschrift oder Rechnung)
- persönliche Gestaltung der Trefferliste
- persönliche Zusammenstellung von Datenbanken
- persönlicher Pushdienst / Suchprofile
- persönliche Kontoeinsicht
- persönliche Informationen zu Workshops, Messen, Veranstaltungen
- Mengenrabatt ab Mindestnutzung von 250 Euro/Monat

Am Beispiel des Newsarchives und GBI kann man die momentanen Probleme bei „content charging“ sehr gut deutlich machen. Man hat eine klare Einteilung in zwei Arten von Benutzern. Für die registrierten Benutzer also vernehmlich Großkunden wie z.B. Bibliotheken ist es ein Fortschritt, da ihre prioren Anforderungen Qualität und Service sind außerdem können Sie bequem per Rechnung am Ende des Monats zahlen. Für den nicht-registrierte Nutzer allerdings hat sich das Informationsmedium Internet klar verschlechtert da z.B. das Zahlen ausschließlich mit Kreditkarte einfach zu unflexibel. (zumal es kein Gesetz gibt, dass man eine Kreditkarte besitzen muss)

Kostenlos vs. Kostenpflichtig

Auf Grundlage der angeführten Schwächen kostenloser Dienste und am Beispiel von „oneShell“ stellvertretend für alle kostenpflichtigen Dienste lässt sich folgendes Fazit ziehen.: Kostenlose Dienste können nicht die Qualität und Zuverlässigkeit von kostenpflichtigen Diensten erbringen. Allein der Einsatz der Technik lässt sich nicht durch Werbung finanzieren. (Beispielsweise liegen gängige Preise für Werbebanner im Bereich von 5,- Euro monatlich. Zum Vergleich: 1000 MByte Webspace kosten bei „oneShell“ 30,-Euro, die Durchschnittspreise für Server liegen bei 80,-Euro.)

Eine solche Serviceleistung durch Fördermittel zu finanzieren ist schon deshalb unmöglich, weil Förderungen prinzipiell nur die Zeit überbrücken bis sich ein Unternehmen/Service selber finanziert.

Kostenpflichtige Serviceleistungen werden kostenlose S. allerdings nicht komplett aus dem Markt verdrängen können. Als Grund sehe ich hierfür im Wesentlichen, dass die hohe Anforderung an Sicherheit, Verfügbarkeit und zugesicherten Dienstgüteparametern für viele Endverbraucher (vor allem im Privaten Bereich da der Endnutzer hier meist nicht darauf angewiesen ist) nur eine untergeordnete Rolle spielt. Diese Nutzer werden dann eher eine langsame Anbindung und höhere Ausfallwahrscheinlichkeiten in Kauf nehmen und eine vergleichbare Serviceleistung, die kostenlos angeboten wird, wählen.

Nur welche Zahlungsmöglichkeit soll man im Internet bevorzugen?

Für größere Beträge können die Zahlungssysteme aus dem täglichen Leben wie EC oder Kreditkarte Verwendung finden. Doch gerade bei Inhaltbezogenen Serviceleistungen sind Klein- und Kleinstbeträge im Bereich von 0,1 bis 10 Cent gängig. Eine zunehmende Kommerzialisierung des Internets macht also auch die Einführung neuer Zahlungsmodelle erforderlich.

5.3 Zahlungsmodelle

Obwohl geradezu dafür gemacht, sind Zahlungen im Electronic Commerce (E-Commerce) die am wenigsten akzeptierte Erscheinungsform. Aus den bisher über 100 für das offene und unsichere Internet konzipierten Vorschlägen für elektronische Zahlungssysteme hat sich kein einziges in der Praxis durchsetzen können.[6] Das Fehlen einer kontextadäquaten, akzeptierten und hinreichend sicheren Form der Wertübertragung behindert gegenwärtig die Entwicklung elektronischer Märkte im Internet. Sichere und effiziente elektronische Geld- und Zahlungssysteme sind jedoch eine wesentliche Voraussetzung für die Durchführung von Transaktionen im Internet. Die Mehrheit der aktuellen Zahlungen im Internet erfolgt immer noch auf dem Kreditkartenweg und erfordert so die Preisgabe von sensiblen Daten an ein sehr unsicheres Medium. Nachrichten über Sicherheitslücken bei Onlinebanken, das Fälschen von Überweisungen im Onlinebanking oder die Betrugsstatistiken der Kreditkartenunternehmen tragen zusätzlich zur beobachtbaren Zurückhaltung bei der Nutzung elektronischer Zahlungssysteme bei.

Die Potenziale der Informations- und Kommunikationstechnik für elektronisches Bezahlen sind zwar technisch fortgeschritten, aber bezüglich der Akzeptabilität nicht ausgeschöpft.

So zeigt sich für die Zahlungsabwicklung ein Beharrungsvermögen zugunsten traditioneller Methoden, wie etwa das Bezahlen per Vorkasse, per Nachnahme oder per Rechnung. Kennzeichnend für dieses Verhalten der Nutzer ist das Vorgehen beim Bezahlen im Internet: Zumeist müssen sie aus den ihnen zur Verfügung stehenden Zahlungssystemen ein vorgegebenes auswählen, ohne sich dabei über die Auswirkungen dieser Wahl auf ihre Sicherheits-, Funktionalitäts- und Wirtschaftlichkeitsanforderungen im Klaren zu sein. Letztlich müssen sich die Nutzer zur Abwicklung einer Transaktion im gegenwärtigen Zustand immer „blind“ zugunsten eines der Zahlungssysteme entscheiden, das der Anbieter vorsieht. nach [6] gehen die Nutzer damit das Risiko ein, dass sich ihre Anforderungen an die Sicherheit, Funktionalität und Wirtschaftlichkeit der Zahlungsdurchführung in einer bestimmten Transaktionssituation nicht erfüllen.

5.3.1 Anforderungen

Die Anforderungen können grob in zwei Gruppen unterteilt werden. Zur ersten, die allgemein unter Sicherheitsanforderungen zusammengefaßt werden können, gehören die Authentizität, die Verbindlichkeit, die Vertraulichkeit und die Integrität. In die zweite Gruppe kommen jene Anforderungen, die ein Zahlungsmittel charakterisieren. Dazu gehören Zuverlässigkeit, Anonymität, Flexibilität, Konvertierbarkeit, Leistungsfähigkeit, Geschwindigkeit beim Einsatz, Akzeptanz des Anbieters und die Leichtigkeit der Integration und Benutzung. Da Anforderungen in dieser zweiten Gruppe nicht voll vereinbar sind, muß jeder Einzelne gemäß den eigenen Prioritäten entscheiden, welches Zahlungssystem den eigenen Bedürfnissen am besten entspricht.[16]

Authentizität

Hier ist die Authentizität des Zugreifers und die der Information zu unterscheiden. Dabei ist erstere als eindeutige Identifizierung zu verstehen, um einen Daten- oder Dienstleistungszugriff zu autorisieren oder zu verweigern. Bei Authentizität der Information muß sich der Benutzer sicher sein, daß diese vom gewünschten Anbieter stammt und nicht durch einen Dritten verfälscht wurde.[16]

Verbindlichkeit

Die Verbindlichkeit ist vor allem im Handel ein wichtiger Faktor, da die Geschäftsparteien in einem Streitfall in der Lage sein müssen, die bereits vergebenen Aufträge bzw. erbrachten Leistungen nachweisen zu können.[16]

Vertraulichkeit

Vertrauliche Daten müssen stets als solche behandelt werden. Daher sind diese bei der Übertragung über ein öffentliches Netz verschlüsselt zu übermitteln. Der Idealfall würde

darin bestehen, daß niemand außer Adressat und Adressant von der Existenz dieser Daten Kenntnis hätten.[16]

Integrität

Man spricht von der Integrität von Daten, wenn diese nicht manipuliert wurden. Jeder Versuch von Unbefugten, Daten während einer Übertragung, zu entziffern, müssen eine Kennzeichnung der Integritätsverletzung zur Folge haben.[16]

Zuverlässigkeit

Mit der Zunahme des Geldverkehrs im Internet muß die Zuverlässigkeit der Zahlung absolut gewährleistet sein. Die Geldtransaktionen und der Geldfluß dürfen weder durch Rowdys noch durch Entwurfsfehler des Zahlungssystems unterbrochen werden oder zum Erliegen kommen.[16]

Anonymität

Für gewisse Geldtransaktionen sollte die Identität der Geschäftsparteien geschützt werden. Weiterhin sollte es nicht möglich sein das Kaufverhalten zu überwachen oder die Einkommenshöhe und -quelle zu bestimmen. Dies könnte man dadurch realisieren, daß die Kosten der Rückverfolgung einer bestimmten Transaktion den tatsächlichen Nutzen bei weiten übersteigen.[16]

Flexibilität

Man spricht von der Flexibilität eines Zahlungssystems, wenn es mehrere alternative Zahlungsmöglichkeiten anbietet. Diese sind notwendig, da man abhängig von der Transaktion unterschiedliche Bedingungen an diese stellt. D.h. man benötigt Zahlungsgarantien, Mindest- oder Maximalbeträge dürfen nicht unter- bzw. überschritten oder es werden Zeitpunktabhängige Zahlung gefordert.[16]

Konvertierbarkeit

Die Anzahl der Anbieter von elektronischen Zahlungssystemen wird mit steigender Benutzeranzahl immer größer werden und somit wird es bestimmt unmöglich sein, daß ein Zahlungsempfänger über die ganze Vielfalt von Zahlungssystemen der unterschiedlichen Anbieter verfügt. Es ist daher wichtig, daß Finanzmittel eines Systems sich leicht in solche eines anderen konvertieren lassen, um den Finanzfluß nicht zu unterbrechen oder zu verlangsamen. Nachdem Geldtransaktionen nicht nur auf das Internet begrenzt bleiben werden, muß diese Konvertierbarkeit auch zwischen realen und virtuellen Geld gewährleistet sein.[16]

Leistungsfähigkeit

Die Tantieme auf den Zugriff von Informationen wird Zahlungen mit kleinen Beträgen zur Folge haben. Somit ist es notwendig, daß durch diese kleinen Transaktionen kein Leistungsabfall im Zahlungsverkehr zu verzeichnen ist. Die Gebühren für solche Minibeträge müssen im Verhältnis zur eigentlichen Geldtransaktion bedeutend kleiner sein.[16]

Geschwindigkeit beim Einsatz

Wie lange eine Transaktion dauern wird, wird zumindest anteilmäßig den Preis beeinflussen. In die Preisberechnung kommen zusätzlich zu den Gebühren eines Zahlungssystemanbieters, auch Telefongebühren, Gebühren für online Dienste, etc.[16]

Akzeptanz des Anbieters

Akzeptanz des Anbieters bedeutet, daß bei einem Händler mit einem bestimmten System bezahlt werden kann. Natürlich wird sich der Einsatz eines elektronischen Zahlungssystems erst dann durchsetzen können, wenn auch genügend Stellen vorhanden sind, die dieses akzeptieren. Diese Anforderung könnte daher über die oben angeführten Anforderungen bestimmen.[16]

Leichtigkeit bei Integration

Benutzer, die mit einem solchen Zahlungssystem eine Transaktion tätigen wollen, sollten Applikationen vorfinden, die diese Zahlungsinfrastruktur beinhalten. Im Idealfall würde diese durch eine vom Zahlungssystem unabhängige Schnittstelle realisiert werden.[16]

Leichtigkeit der Benutzung

Benutzer sollten nicht fortwährend unterbrochen werden, um mit Zahlungsinformationen versorgt zu werden. Zahlungen bis zu einem vom Benutzer festgelegten Betrag sollten automatisch erfolgen, andere die das festgelegte Limit übersteigen sollten ausdrücklich genehmigt werden. Die Benutzer müßten jedoch jederzeit in der Lage sein, ihre Gesamtausgaben zu begrenzen.[16]

5.3.2 Klassifikation Digitaler Zahlungssysteme

Um Bezug auf die Risiken der Verwendung digitaler Zahlungsmodelle, Gemeinsamkeiten verschiedener Systeme identifizieren zu können, werden die digitalen Zahlungssysteme klassifiziert.

1. [10] [Janson und Waidner,1996] definieren anhand des Kriteriums

Zeitpunkt der Zahlung drei unterschiedliche Arten von Zahlungssystemen:

- Prepaid-Zahlungssysteme mit der Belastung vor der Zahlung.
Bei Prepaid-Systemen muß der Kunde, bevor er eine Zahlung ausführt, ein Guthaben auf sein Zahlungsmedium einzahlen, d.h. es entsteht eine gewisse Zeitspanne zwischen dem Einzahlen des Geldes und der Ausgabe (z.B. Geldkarte, Telefonkarten, Electronic Cash). Das Geld wird zuerst von einem Bankkonto abgebucht und in eine zahlungsgerechte Form gebracht. Das so vorbereitete Geld wird später zum Konsum verwendet. Die digitale Geldbörse (Geldkarte) des deutschen Kreditwesens ist ein Beispiel für Prepaid-Systeme.
- Pay now-Zahlungssysteme mit der Belastung während der Zahlung.
Mit dem Auslösen einer Zahlung wird sofort die entsprechende Belastung auf dem Bankkonto des Kunden ausgeführt, d.h. es ist keine „Zwischenlagerung“ des Geldes nötig. Ein Beispiel für ein pay-now System ist EC-Direct mit der EC-Karte als POS/EFT System. Konsum und Abbuchung vom Konto erfolgen (im Prinzip) zum selben Zeitpunkt.
- Pay later-Zahlungssysteme mit der Belastung nach der Zahlung.
Die Zahlung ist genau gesehen ein Zahlungsanweisung, da erst nach einem bestimmten Zeitintervall oder nach der Kumulierung von Beträgen die Abbuchung auf dem Bankkonto des Kunden erfolgt. Hier erfolgt der Konsum vor der Abbuchung vom Bankkonto; es entsteht ein Kredit, der durch den Verkäufer oder die vermittelnde Organisation finanziert wird.
Kreditkarten- und Schecksysteme sind typische Pay later-Systeme.
Der Zahlungszeitpunkt bezieht sich dabei auf die Zeit, die zwischen dem Auslösen einer Zahlungsanweisung und der tatsächlichen Belastung auf dem Kundenkonto liegt. Für Prepaid-Systeme wird in dieser Arbeit auch der Ausdruck „bargeldähnlich“ verwendet.
- **Zahlung im Voraus auf ein Guthabekonto.** Der Kunde zahlt einen bestimmten Betrag im voraus auf ein Guthabekonto ein und bekommt solange Waren ohne weitere Zahlung geliefert, wie deren Wert den vorausbezahlten Betrag auf dem Konto nicht überschreitet. Erst dann fordert der Händler eine neue Zahlung an. Ist der Handel mit einem Händler komplett, wird das Restguthaben zurückgezahlt. Hauptvorteil dieses Verfahrens ist, daß damit der Zeitaufwand durch die vielen unter Umständen sonst notwendig werdenden kleinen Einzelzahlungen eingespart werden kann.
- **Zahlung je Zeiteinheit.** In diesem Fall zahlt der Kunde nicht für ein Stück Ware, sondern für eine bestimmte Zeit einer Dienstleistung, etwa für die Nutzung eines Online-Spiels. Nach Ablauf der bezahlten Zeitspanne bekommt der Kunde automatisch erneut den Betrag für die nächste Zeitspanne in Rechnung gestellt.

Pay-now- und Pay-later-Systeme lassen sich der Klasse der kontobasierten Systeme zuordnen.

2. **bei der Bindung an eine Kontoführung** unterscheidet man digitale Zahlungsmittel anhand der Existenz einer im Hintergrund mitlaufende Kontoführung.
 - Notational money-Zahlungssysteme (Kontobasierte Zahlungssysteme).
Geld existiert lediglich auf Konten. Es wird durch Verfügung des Zahlenden (vgl. Überweisung) oder des Zahlungsempfängers (vgl. Lastschriftinzug) vom Konto des Zahlenden auf das Konto des Zahlungsempfängers übertragen. Kreditkartenzahlungen mit SET sind ein Beispiel für ein Notational money-Zahlungssystem.
 - Token money- oder Stored Value-Zahlungssysteme (Geldähnliche Zahlungssysteme).
Geld existiert in Form von digitalen Werteinheiten, die ähnlich Bargeld den Besitzer wechseln können, ohne daß eine Kontoübertragung damit verbunden ist. Als Beispiel hierfür läßt sich etwa Mondex anführen.
3. **bei der Art der Zahlungskommunikation** unterscheidet man, ob der Zahlende direkt oder indirekt mit dem Zahlungsempfänger in Verbindung tritt.
 - „Online“-Zahlungssysteme.
Der Zahlende tritt beim Zahlungsvorgang direkt (online) mit dem Zahlungsempfänger in Verbindung. Kreditkartenzahlungen mit SET werden in direkter Kommunikation zwischen Zahlendem und Zahlungsempfänger ausgeführt.
 - „Offline“-Zahlungssysteme.
Der Zahlende tritt beim Zahlungsvorgang indirekt (offline) mit dem Zahlungsempfänger in Verbindung. Bei einer Zahlung mittels Überweisung im Online-Banking tritt der Zahlende nicht direkt mit dem Zahlungsempfänger, sondern mit seinem Kreditinstitut in Verbindung.
4. **nach dem Initiator eines Zahlungsvorgangs** Abhängig davon, wie bzw. von wem ein Zahlungsvorgang initiiert wird, kann man unterscheiden in:
 - Push-Zahlungssysteme.
Der Zahlende initiiert den Zahlungsvorgang, indem er den zu zahlenden Betrag über die eingebundenen Zahlungssystembetreiber (vom Bankensystem des Zahlenden bis zum Bankensystem des Zahlungsempfängers) auf ein Konto des Zahlungsempfängers überweist.
Ein Beispiel für ein Push-Zahlungssystem bei kontobasierten Zahlungssystemen ist der Einsatz von Online-Banking unter Einsatz von HBCI. Auch geldähnliche Zahlungssysteme fallen in diese Rubrik, bei denen der Zahlende den zu zahlenden Betrag in Form digitaler Werteinheiten zum Zahlungsempfänger überträgt, ohne daß im Hintergrund Zahlungssystem-Betreiber involviert sein müssen, z.B. Zahlung mit der Mondex-Karte.
 - Pull-Zahlungssysteme.
Der Zahlungsempfänger initiiert den Zahlungsvorgang, indem er seine Bank anweist, den vereinbarten Betrag über die Bank des Zahlenden vom Konto des Zahlenden einzuziehen. Das elektronische Lastschriftverfahren ist ein

Beispiel für ein „Pull“-Zahlungssystem. Der Zahlende gibt dabei dem Zahlungsempfänger die Ermächtigung (etwa in Form einer Lastschrift einzugsermächtigung, eines Schecks oder eines Kreditkartenbeleges), über die beteiligten Zahlungssystem-Betreiber das Geld einzuziehen.

Aus diesen Zahlungssystemklassen wird ein generisches Modell abgeleitet. Dieses Modell der generischen Zahlungssystemklassen ist gleichzeitig die Grundlage für die Risikoanalyse der digitalen Zahlungssysteme, die sich an den Informationsströmen orientiert und auf den Kriterien der mehrseitigen Sicherheit basiert.[7]

Generische Zahlungssystemklassen

In [Abad-Peiro, Asokan, Steiner, Waidner 1996] werden die Unterscheidungsmerkmale „Art der Zahlungskommunikation“ „Bindung an Kontoführung“ sowie „Initiierung des Zahlungsvorganges“ gemäß ihrer Informationsströme kombiniert. Dieser Ansatz wird hier als geeigneter Anknüpfungspunkt für die Risikoanalyse verwendet. Insbesondere wird die Einteilung in die nachfolgenden „Generischen Zahlungssystemklassen“ übernommen:

1. „Geldähnliche“ / „Cash-like“-Zahlungssysteme (online). Die zwischen Zahlendem und Zahlungsempfänger übertragenen digitalen Informationen stellen selbst einen Wert dar (Beispiel: eCash von DigiCash).

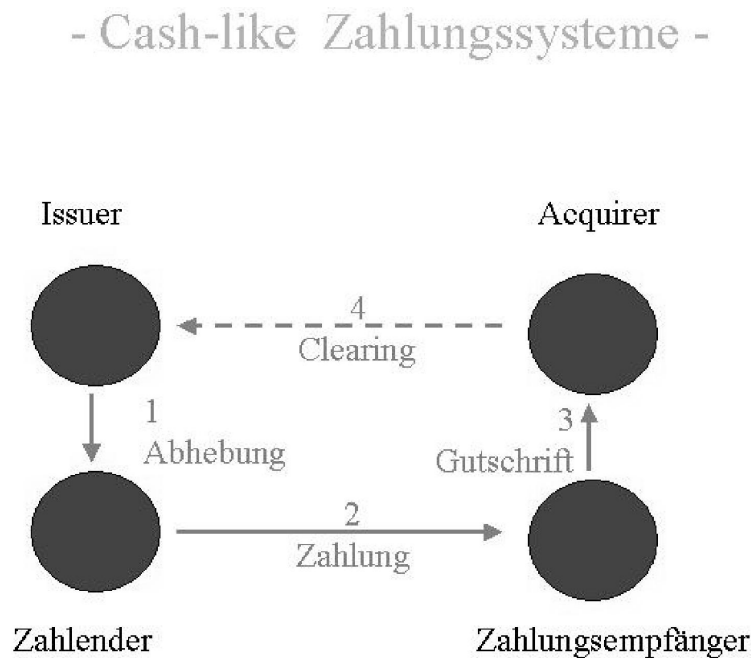


Abbildung 5.1: Geldähnliche Zahlungssysteme

2. „Scheckähnliche“ / „Cheque-like“-Zahlungssysteme (online). Der Zahlende ermächtigt den Zahlungsempfänger durch Übergabe eines Inhaberpapiers (Beispiel: eCheck von FSTC) oder der Übermittlung seiner Kreditkarteninformationen (Beispiel: SET) dazu, den vereinbarten Betrag über das Bankensystem von seinem Konto einzuziehen. Der Begriff „Scheckähnliche Zahlungssysteme“ wird hier anstelle des Begriffes „Notational money“ oder „Kontobasierte“ Zahlungssysteme verwendet, weil sonst geldähnliche Zahlungssysteme wie eCash, bei denen ebenfalls Konten zur Abwicklung einer Zahlung involviert sind, nicht eindeutig einer Klasse zuzuordnen wären.

- Cheque-like Zahlungssysteme -

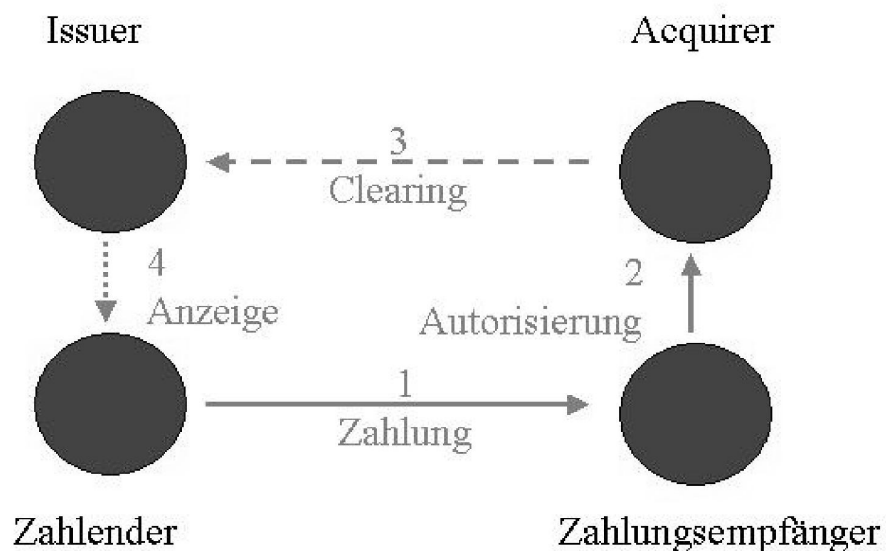


Abbildung 5.2: Scheckähnliche Zahlungssysteme

3. „Push-like“-Zahlungssysteme (offline). Der Zahlende weist seine Bank an, von seinem Konto einen bestimmten Betrag auf das Konto des Zahlungsempfängers zu überweisen (Beispiel: Online-Banking mit Überweisungsauftrag).

- Push-like Zahlungssysteme -

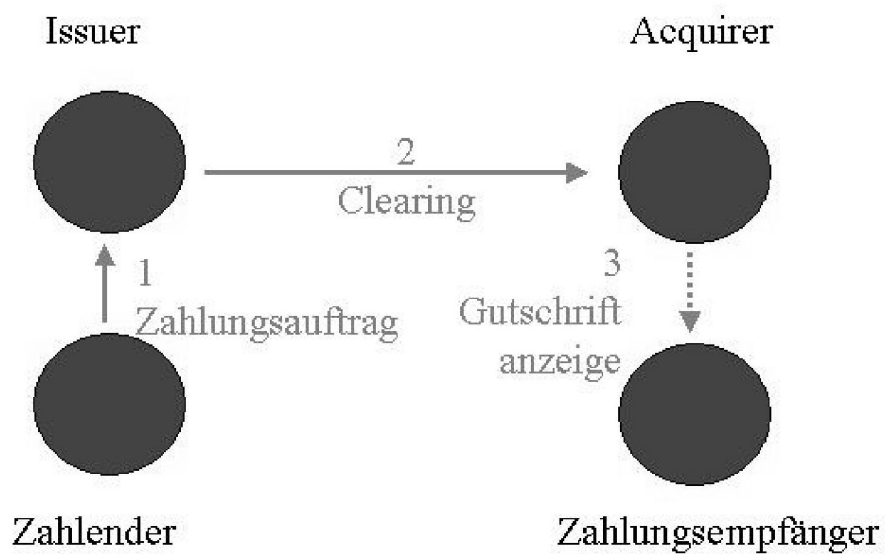


Abbildung 5.3: Push-like Zahlungssysteme

4. „Pull-like“-Zahlungssysteme (offline). Der Zahlungsempfänger weist seine Bank an, den vereinbarten Betrag über die Bank des Zahlenden vom Konto des Zahlenden einzuziehen (Beispiel: Elektronisches Lastschriftverfahren).

- Pull-like Zahlungssysteme -

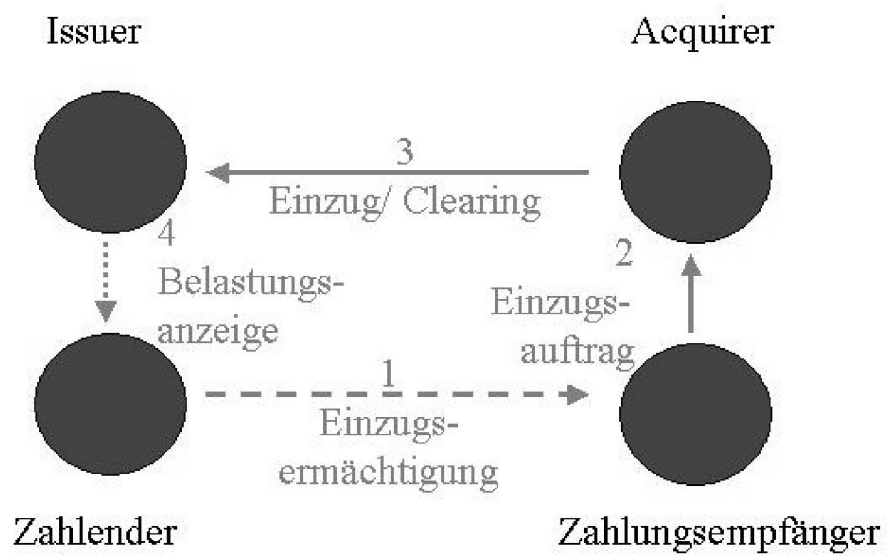


Abbildung 5.4: Pull-like Zahlungssysteme

Auch die Unterscheidung nach dem Zeitpunkt der Zahlung läßt sich in dieses Modell einarbeiten. Die „geldähnlichen“ Online-Zahlungssysteme entsprechen i.d.R. dem Typ der „Prepaid“-Zahlungssysteme. Die „Push- und Pull-like“-Offline-Zahlungssysteme entsprechen i.d.R. dem Typ der „Paylater“-Zahlungssysteme. Die „scheckähnlichen“ Online-Zahlungssysteme können im Falle von ec-Cash (POS) mit der EC-Karte dem Typ der „Pay now“-Zahlungssysteme entsprechen, können aber im Falle von Kreditkartenzahlungen auch den „Pay later“-Zahlungssystemen zuzurechnen sein.

5.3.3 Zahlungsmodelle

1. **Electronic currency (Elektronische Währung)** Beim Modell der elektronischen Währung kauft der Benutzer elektronische Währungszertifikate von einem Währungsserver. Er zahlt diese Zertifikate durch die Abrechnung von einem zuvor eingerichteten Konto auf dem Währungsservers, durch eine Kreditkartenzahlung, elektronische Schecks oder reales Geld, welches z.B. über einen umgekehrten Geldautomaten angenommen werden kann. Findet eine elektronischen Währungsemission durch den Währungsserver an den Benutzer statt, dann entspricht der Betrag dieser elektronischen Währung den gleichen Betrag, der vorher eingezahlt wurde. Eine Transaktion mit elektronischen Geld kann dann bei einer Privatperson, einem Händler oder einem anderen Dienstleistungsunternehmen getätigt werden. Die Geschäftspartner lassen ihrerseits die Zertifikate auf ihr Konto gutschreiben.[8]

Anforderungen: Eine elektronische Münze ist nichts anderes als digitale Information, welche beliebig kopiert werden kann. Damit diese ihre Funktion als Zahlungsmittel wahrnehmen kann, müssen verschiedene Anforderungen erfüllt sein:

- Fälschungssicherheit muss bei elektronischen Münzen durch spezielle Vorkehrungen erreicht werden, während bei Banknoten Fälschungen durch physikalische Merkmale (Wasserzeichen, Metallfaden, Spezialpapier, etc.) verhindert werden. Es muss möglich sein, eine ungültige Münze effizient zu erkennen.[8]
- Einmalige Verwendbarkeit Mehrfaches Ausgeben derselben Münze (Double-Spending) ist bei digitalem Geld ohne spezielle Vorkehrungen problemlos möglich. Grundsätzlich kann dies nicht verhindert, sondern nur detektiert werden. Double-Spending kann sowohl vom Kunden bei der Bezahlung, wie auch vom Verkäufer beim Einzahlen ins Bankensystem, vorkommen.[8]
- Anonymität wird dadurch erreicht, dass Zahlungen nicht nachvollzogen werden können, dies aus Datenschutzgründen. Es soll verhindert werden, dass bei der Einzahlung festgestellt werden kann, welcher Kunde die Münze abgehoben hat.[8]
- Schutz gegen Kriminalität Elektronisches Geld kann vollständig anonym gehalten werden. Dies ist aus Sicht der Kunden, welche Wert auf Datenschutz legen, eine legitime Forderung. Gleichzeitig aber wird durch die Anonymität ein großes Potential für Kriminalität geschaffen. So bietet elektronisches Geld enorme Möglichkeiten zur Geldwäscherei oder auch zur perfekten Erpressung (das Geld kann z. B. in einer Zeitung abgedruckt werden). Es ist daher wünschenswert, dass der Staat oder eine andere, berechnigte Instanz auf richterlichen Befehl hin die Anonymität einer Münze aufheben kann.

[11] NetCash und e-Cash haben dieses Modell zur Grundlage. Es ist für Geldtransaktionen von \$ 0.01 bis \$ 10 konzipiert. Hauptvorteil dieses Modells ist das große Potential der Anonymität. Größter Nachteil des elektronischen Währungsmodell ist die Notwendigkeit, riesige Datenbanken zu unterhalten, die bereits getätigte Geldtransaktionen speichern, um doppelte Ausgaben zu verhindern. Dabei speichert man entweder die Zertifikatsnummern, die bei der Zertifikatsemission registriert wurden und noch im Umlauf sind oder jene, die bei der Zertifikatsemission nicht registriert wurden aber eingegangen sind.

2. **Credit-debit instruments (Soll- und Habenmodell)**[8] Im Zahlungssystem, die auf dem Soll- und Habenmodell basieren, müssen die Benutzer zunächst auf einem Zahlungsserver registriert werden. Dabei erhält jeder ein Konto und die Bevollmächtigung dieses belasten zu können. Das Benutzerkonto wird bei der Soll- oder Scheckmethode zusätzlich zu dem Betrag der Transaktion oder dem eingereichten Scheck mit Gebühren belastet. Bei der Habenmethode werden die Gebühren auf der Habenseite des Kontos verbucht, diese werden dem Benutzer dann in Rechnung gestellt oder bilden ein Saldo eines Kundenkontos, zu dessen Ausgleich er aufgefordert wird. Die Implementierung für diese elektronisches Zahlungssystem ist für beide Methoden gleich.

Repräsentanten dieses Modells sind NetBill und NetCheque. Es ist für Geldtransaktionen von \$ 10 bis über \$ 1000 konzipiert. Der Vorteil des Soll- und Habenmodells liegt in der Möglichkeit die Zahlung zu überprüfen, d.h. ob diese geleistet worden sind und wer der Indossant ist. Diese Art von Zahlungen sind in der Geschäftswelt äußerst wichtig und bilden bei Privatpersonen einen großen Anteil ihrer Geldtransaktionen. Dieses System schützt zur Zeit nicht besonders gut die Anonymität. Hierin ist der große Nachteil zu sehen.[8]

3. **Secure credit card presentation (Sichere Kreditkarten)** Dem Zahlungssystem, welches man die größten Chancen einräumt sich im Markt durchzusetzen bilden die Geldtransaktionen mit Kreditkarten. Dabei wird die Kreditkartennummer des Benutzers mit einem öffentlichen Schlüssel chiffriert, so daß sie nur vom Geschäftspartner oder einer Stelle, die den Zahlungsverkehr abwickelt, decodiert werden kann. Bei diesem Zahlungssystem müssen die Benutzer zunächst nicht registriert werden, sondern benötigen lediglich eine Kreditkartennummer. Dieser Umstand bildet die Grundlage für eine breite Akzeptanz seitens der Geschäftsstellen. Die Verschlüsselung der Kreditkartennummer verhindert Lauschangriffe vom Netz. Bei einer Codierung von der Kreditkartennummer und dem zu zahlenden Betrag mit einem Schlüssel eines Finanzunternehmens, die den Zahlungsverkehr abwickelt, kann der Händler die Kreditkartennummer seines Kunden nicht mehr dechiffrieren. Diese Maßnahme schützt den Kunden zusätzlich vor Mißbrauch seiner Kreditkarte. Nachteilig zeigen sich jene Systeme, die keine Unterschrift erfordern, da jemand einen Kreditkarteneinsatz durchführen könnte, wenn er die Kreditkarteninformationen kennt. Ein weiterer Nachteil ist der unwahrscheinliche Gebrauch der Karte bei sehr kleinen Beträgen, da die Standardgebühren von Kreditkarteninstituten diese bei weiten übersteigen werden.

Es ist für Geldtransaktionen von \$ 10 bis über \$ 1000 konzipiert. Ein Anbieter dieses Zahlungssystems ist Green Commerce. Hauptvertreter jedoch sind die großen Kreditkarteninstitute, die sich auf ein einheitliches System SET geeinigt haben.

Bei der Betrachtung der Spezifikation wird man jedoch feststellen, daß die Sicherheitssysteme weiter ausgebaut wurden, so daß die Sammlung von Daten über Kreditkarteninformation durch Händler verhindert wird. Eine Datensammlung durch Kreditkarteninstitute, die Kaufverhalten des Kunden protokollieren, wird ebenfalls vereitelt.

4. **Direct transfer and collection agents (Direkte Überweisung und Sammelagenten)** Das direkte Überweisungsmodell ähnelt in gewissen Maßen das Soll- und Habenmodell allerdings müssen sowohl Kunden als auch Händler Konten auf dem Zahlungsserver besitzen. Soll eine Zahlung getätigt werden, ordnet der Kunde dem Server an, eine Überweisung auf das Konto des Händlers zu machen. Dieser wird von der eingehenden Zahlung benachrichtigt und stellt dem Kunden, nach Überprüfung der Transaktion, ein Produkt, eine Dienstleistung und/oder eine Information dem Kunden zur Verfügung.

Bei Sammelagenten tritt der Kunde mit dem Händler in Kontakt und letzterer stellt ihm dann die nötigen Informationen für den Kauf zur Verfügung. Diese Informationen, zusammen mit der Zahlung, sammelt eine dritte Partei, der sogenannte Sammelagent. Anschließend wird dem Händler eine Quittung zugestellt, der dann die Gegenleistung erbringen muß. Diese Methode wird zwar nicht von einem spezifischen Kreditinstitut angeboten, bildet allerdings die idealen Voraussetzungen, um die bereits erwähnten Zahlungssysteme in sich zu vereinigen, so daß ein Händler nicht mehr die ganze Vielfalt von Systemen anbieten muß. [8]

5.3.4 Existierende Systeme

Secure Electronic Transactions (SET)

[8] Die Entwicklung der „Secure Electronic Transaction“ Spezifikation ist das Ergebnis einer Gemeinschaftsarbeit von Visa, MasterCard, IBM, Netscape, Microsoft und CyberCash. Mit SET kann ein Benutzer ein Produkt, Dienstleistung oder Information mit Kreditkarte bezahlen, ohne dabei seinem Geschäftspartner seine Kreditkartennummer zu übermitteln. SET genügt folgenden Sicherheitserwartungen im elektronischen Handel:

- Geheimhaltung der Zahlungs- und Bestellinformationsübermittlung
- Authentizität des Karteninhabers für ein eindeutig bestimmtes Bankkonto (diese Authentizität wird sichergestellt durch den Gebrauch von einer digitalen Unterschrift und Karteninhaber Zertifikate)
- Authentizität des Händlers für die Autorisierung zur Annahme von Kreditkartenzahlungen (diese Authentizität wird sichergestellt durch den Gebrauch von einer digitalen Unterschrift und Händler Zertifikate)
- Zahlungsinformationsintegrität wird durch den Gebrauch von digitalen Unterschrift sichergestellt
- Spezielle Zweckzertifikate

- Zurückweisung zweifelhafter Zahlungsvorgänge

Die Bezahlung kann entweder Online oder per E-Mail vollzogen werden. Der Zahlungsvorgang kann grob in 5 Schritte eingeteilt werden, diese sind:

1. die Registrierung des Kartenbesitzers (Cardholder Registration),
2. die Registrierung des Händlers (Merchant Registration),
3. der Kaufauftrag (Purchase Request),
4. die Zahlungsautorisierung (Payment Authorization) und schließlich
5. die Zahlungsannahme (Payment Capture).

Ziel der Registrierung des Kartenbesitzers ist der Erhalt eines Zertifikats, mit dem er sich beim Händler als autorisierter Besitzer einer Kreditkarte ausweisen kann, ohne dabei diesem Kreditkarteninformationen preiszugeben. Um dieses Zertifikat zu bekommen, muß der Karteninhaber die Registrierung starten und auf die Antwort der CA (Certificate Authority) warten. Wenn er eine Antwort erhält, kann er das Formular für die Registrierung beantragen, das ihm dann durch die CA zugesandt wird. Erst mit diesem kann er das Zertifikat anfordern, welches dann durch die CA erzeugt und anschließend dem Karteninhaber zugeschickt wird. Informationen, die ab dem Zeitpunkt, in dem der Benutzer das Formular für die Registrierung an die CA schickt, werden mit raffinierten kryptographischen Algorithmen verschlüsselt und mit digitalen Unterschriften unterzeichnet. Wenn die CA das Zertifikat dem Karteninhaber zusendet, wurde eine neue Zahl generiert, die aus einer Zufallszahl, der Kreditkartennummer und dem Verfallsdatum der Karte zusammengesetzt ist. Somit werden vertrauliche Information unkenntlich gemacht.

Analog verläuft die Registrierung des Händlers bei der CA. Allerdings wird hierbei eine Verbindung zwischen Händler und Kunden durch die CA geschaffen. Somit stellt der Kunde fest, daß der Händler autorisiert ist, Kreditkartenzahlungen zu akzeptieren. Durch diese Verbindung werden nun Käufer und Händler über das weitere Procedere, sprich die Bestellinformationen und die Zahlungsinstruktionen, direkt verhandeln.

Beim Kaufauftrag sendet der Käufer dem Händler sein Zertifikat und die Bestellinformationen; er erhält im Gegenzug das Händlerzertifikat, die Zahlungsinstruktionen und ein eindeutiger Transaktionsidentifikator. Letzter wird an die Kreditkartengesellschaft, in Wirklichkeit an den Payment-Server, geschickt. Diese wiederum kann mit den Daten des Kreditkarteninhabers und dem Transaktionsidentifikator dem Händler eine Zahlungsnachricht senden. Auch in dieser Phase sind die Informationsübermittlungen mit kryptographischen Algorithmen verschlüsselt und mit digitalen Unterschriften unterzeichnet.

Die Zahlungsautorisierung sieht vor, daß sich der Händler mit einer Deckungsanfrage beim Payment-Server über den Status Quo der eingehende Kreditkartenzahlung erkundigt. Dabei werden unter anderem vom Server der Transaktionsidentifikator, der zu zahlende Betrag, etc. von Händler und Käufer verglichen. Stimmen alle Daten überein, dann erhält der Händler vom der Kreditkartengesellschaft eine Deckungszusage über den ausgestellten Betrag, und später auf sein Konto gutgeschrieben. Dito bezüglich kryptographischen Algorithmen und digitalen Unterschriften.

Bei der Zahlungsannahme bestätigt der Händler dem Käufer den erhaltenen Betrag.

Entscheidender Vorteil dieses Zahlungssystems ist neben der hohen Sicherheit auch der hohe Grad an Anonymität, da die Kreditkartengesellschaft keine Informationen über die gekauften Waren erhält und Händler nicht an die Kreditkarteninformationen gelangen. Diese Technik wird durch die „Duale Unterschriften (dual signatures)“ verwirklicht. Die Bestellinformationen und die Zahlungsinstruktionen sind jedoch so miteinander verknüpft, daß der Händler die Zahlungsbestätigung nur erhalten kann, wenn er sich an die mit dem Kreditkarteninhaber ausgehandelten Bestelldaten hält.

Aus diesen Gründen rechnet man diesem Zahlungssystem für Geldtransaktionen, über ein bestimmten Betrag, im Internet die größten Chancen ein.

Green Commerce

[8] Green Commerce wurde von First Virtual entwickelt, und basiert auf dem Prinzip einer Kreditkartenzahlung ohne dabei vertrauliche Daten über das Netz zu schicken. Der Benutzer kann, wenn er eine E-Mail Adresse besitzt, ein Account bei der First Virtual einrichten. Dabei wird ihm eine Kartennummer, die sogenannte „VirtualPIN“ zugewiesen. Möchte ein Kunde etwas kaufen, so muß er zunächst dem Händler seine VirtualPIN zusenden. Dieser schickt sie zusammen mit der Anfrage, die Kreditkarte des Käufers mit dem zu zahlenden Betrag zu belasten, weiter an First Virtual. Diese wiederum ein E-Mail an den Käufer schickt, um die Zahlung zu bestätigen. Die Rückantwort des Benutzers kann entweder „Ja“ „Nein“ oder „Betrug“ sein. Nur wenn die Antwort „Ja“ lautet, wird die Geldtransaktion durchgeführt.

Kritiker von kryptographischen Algorithmen sind der Meinung, daß Schlüssel zur Chifrierung geknackt werden können. Hinzu kommt, daß sich Viren in den Benutzerrechner einnisten können, um Kreditkarten- oder Kontoinformationen auszuspionieren, die dann unbemerkt ins Netz geleitet werden. Bei Green Commerce können Transaktionen unverschlüsselt im Internet versendet werden, da sie für Dritte, die diesen Informationsaustausch abhören, keine verwertbare Information darstellt. Selbst wenn die VirtualPIN in die Hände eines Unbefugten fällt, kann dieser keine Mißbrauch damit treiben, da erstens dieses Zahlungssystem nur im Internet verwendet werden kann und zweitens ihre Anwendung eine E-Mail Bestätigung erfordert. Ein weiterer Vorteil liegt in der Möglichkeit die Zahlung Off-Line zu tätigen.

NetBill

[8] NetBill ist eine Entwicklung eines Zahlungssystems von der Carnegie Mellon University, welches hauptsächlich im Bereich der „Micropayments“ eingesetzt werden soll, d.h. für den Handel von elektronischen Daten und Informationen, deren Wert nur Bruchteile eines Dollars darstellen. Bei einer Transaktion schickt zunächst der Händler dem Käufer die geforderte Ware verschlüsselt und unverschlüsselt die Zahlungsinstruktionen. Der Käufer der die Information erhält, sendet dem NetBill-Server einen Produktidentifikator und dem Händler eine „Electronic Payment Order“, die der Händler, nachdem er sie bekommt,

zusammen mit dem eigenen Produktidentifikator weiter an den NetBill-Server schickt. Stimmen die Produktidentifikatoren überein, wird das Konto des Käufers mit dem zu zahlenden Betrag belastet und dem Konto des Händlers gutgeschrieben. Es folgt eine Bestätigung über die getätigte Transaktion an den Händler, der dann schließlich den Schlüssel zur Dechiffrierung der Information dem Käufer zuschickt.

NetCheque/NetCash

[8] Sowohl NetCheque als auch NetCash wurde vom Wissenschaftsinstitut für Information der University of Southern California entwickelt. Den NetCheque kann man sich als einen digitalen Papierscheck vorstellen. Für den Benutzer dieses Zahlungssystems wird ein Konto auf einem Kontoserver eingerichtet, von welchem aus er einen Scheck anfordert. Dieser wird mit dem Namen des Zahlungsempfänger und dem des Käufers, seiner Bank, seiner Kontonummer und mit dem zu zahlenden Betrag versehen, an den Händler zugesendet. Der Händler indossiert dann den Scheck und schickt ihn an seinen Kontoserver. Der Kontoserver des Käufers belastet sein Konto mit dem zu zahlenden Betrag und schreibt es dem Konto des Händlers gut.

Nachdem NetCheque die Anonymität von Kunde und Händler nicht schützt, wurde NetCash hinzugezogen. Ein System bei dem die Identität der Geschäftsparteien bis zu einem gewissen Grad im verborgenen bleibt, indem mit elektronischen Geld gezahlt wird. Dabei werden die NetCheques gegen digitales Geld eingetauscht, mit dem dann gezahlt werden kann. Voraussetzung für den Schutz der Anonymität ist, daß während der Ausgabe der Scheine nicht gespeichert wird, wer sie erhält. Die vom Geldserver ausgegebener Scheine (in der Originalspezifikation wird zwar von „coins“, also Münzen gesprochen, allerdings bevorzuge ich von Scheinen statt von Münzen zu sprechen, da zumindest reale Münzen keine Seriennummer besitzen) werden durch den Server unterzeichnet, und kommen dann auf eine Liste mit den im Umlauf befindliche Scheine. Der Einzahler dieser Scheine bekommt durch den Geldserver, der diese vorher auf ihre Gültigkeit hin überprüft, den Gegenwert auf sein Konto gutgeschrieben.

E-Cash

[8] E-Cash wurde von DigiCash entwickelt, und ist ein Zahlungssystem, welches ausschließlich elektronisches Geld anbietet. Wenn ein Benutzer digitales Geld von seiner Bank verlangt, dann wird zunächst sein Echtgeldkonto mit dem Gegenwert des angeforderten elektronischen Geldes belastet. Es werden ihm auf die Festplatte seines Computers die Scheine mit den angeforderten Beträgen gespeichert, indem sie in ein elektronisches Portemonnaie gelegt werden. Jeder Schein trägt eine Seriennummer, die vorher durch ein Zufallsverfahren erzeugt und dann unsichtbar gemacht wurde. Die Bank speichert nicht, welche Scheine emittiert wurden und an wen sie die Scheine ausgegeben hat. Wenn der Benutzer den Verblendungsfaktor entfernt, erscheint dann wieder die Seriennummer auf den Scheinen, die dann ausgegeben werden können. Kauft der Benutzer ein, werden die Scheine aus seinem Portemonnaie entfernt, von seiner Festplatte gelöscht und an den Händler gesendet. Diese werden anschließend vom Händler an die Bank geschickt, die

dann überprüft, ob die Seriennummern dieser Scheine in der Datenbank der bereits eingereichten Scheine auftauchen. Ist dies der Fall wird der Händler über die Ungültigkeit der Scheine benachrichtigt. Andernfalls wird dem Händler der Gegenwert des Betrages auf sein Echtgeldkonto gutgeschrieben.[8]

5.3.5 Anbieter von Zahlungssystemen

Anhand der Erklärungen zu den verschiedenen Zahlungssystemen muss klar werden, dass es nicht „Das richtige Zahlungssystem“ geben kann da ein System welches sich zum Zahlen von Kleinstbeträgen eignet keinesfalls zum Zahlen von konkreter Ware benutzt werden kann.

Aus diesem Grund haben sich verschiedene Anbieter gebildet, die eine Kombination der verschiedenen Systeme anbieten. (welche natürlich zu dem Bereich der kostenpflichtigen Serviceleistungen gehören)

T-pay

„T-Pay ist die Zukunft des Bezahls im Internet“ -so wirbt zumindestens die Telekom auf www.t-pay.de. T-Pay bietet vier Zahlungsarten für Einkäufe im Internet.

- über die Telekom-Rechnung
T-Pay bietet die Möglichkeit, Einkäufe im Internet über die Telekom-Rechnung zu bezahlen. Voraussetzung hierfür ist, dass man sich vorher registrieren lässt.(Name, E-Mailadresse und Passwort)
Als registrierter Nutzer kann man mit Nutzernamen und Passwort einkaufen - die Zahlung erfolgt auf der monatlichen Telekom-Rechnung.
- mit MicroMoney der Guthabekarte von T-Pay Zum Bezahlen kleinerer Beträge (Größenordnung 1-5 EURO) bietet T-pay diese Möglichkeit an.
Die Karte enthält eine verdeckte 16stellige MicroMoney Nummer, mit der man online bezahlen kann. Vorteil hier ist die Anonymität da man sich nicht registrieren muss. Als Nachteil würde ich den fixen Betrag von 15 EURO bewerten.
- per Lastschrift oder Kreditkarte Hier wirbt T-pay damit, dass registrierte Benutzer nur ihr individuelles Passwort eingeben müssen und dann die Bezahlung mittels Lastschrift oder Kreditkarte durchführen können ohne wiederholt ihre persönlichen Daten angeben zu müssen. Ich glaube nicht, dass sich dies durchsetzen wird, da dies ja ein Speichern dieser Daten an zentraler Stelle voraussetzt.(wo sie natürlich für Angriffe lohnende Ziele darstellen)

The Blue Window (Swisscom)

The Blue Window, kurz Bluewin, ist der Name für die Internet Dienste der Swisscom. Bluewin bietet sowohl Connectivity Dienste für Privatkunden und Firmen wie auch Web-Hosting und Consulting. Ein spezieller Dienst stellt dabei die integrierte Gesamtlösung CommerceMaker dar. CommerceMaker ist eine Transaktionsplattform, welche sicheres Einkaufen und Bezahlen via Internet ermöglicht. Insbesondere besteht die Möglichkeit, Mikrotransaktionen via Kreditkarte oder Swisscom-Telefonrechnung zu belasten. CommerceMaker übernimmt den kompletten Zahlungsverkehr. Einzig beim Kauf auf Rechnung ist der Verkäufer für das Inkasso verantwortlich.

Zahlungsarten:

- Kreditkarte: VISA, MasterCard, American Express, Diners
- Swisscom Telefonrechnung
- Postkonto
- Rechnung
- Wallet

Gesichert wird die Übertragung der Daten via SSL (mit 128 Bit Schlüssel).

Die Installation von CommerceMaker kostet je nach Optionen in der Grössenordnung von Fr. 4500.- bis Fr. 15'000.-. Darin enthalten ist die Lizenz für die Software, das Einrichten sowie ein Tag Support. Weiter fallen monatliche Kosten zwischen Fr. 250.- und Fr. 350.- an. Die Transaktionen sind je nach verwendeter Zahlungsart unterschiedlich. Bei Kreditkarten wird eine Kommission von 2% zusätzlich zu den Gebühren der Kreditkartenfirma verrechnet. Dasselbe gilt für Rechnungen. Mikrotransaktionen via Wallet werden mit 6% Kommission belastet. (Das „Wallet“ ist ein elektronisches Portemonnaie, von welchem Mikrotransaktionen abgebucht werden.)[11]

5.3.6 Analyse potenzieller Modelle auf Akzeptanz

In Deutschland waren 1997 11 Mio Kreditkarten, 40 Mio ec-Karten, 59.000 Terminals in Benutzung. 85 % aller Zahlungsvorgänge wurden noch in Bar getätigt. Als Gründe hierfür sind vorrangig:

- Transaktionsgebühren zu hoch
- Angst vor Mißbrauch
- Anonymität bei Bargeld besser
- teilweise zu wenige Akzeptanzstellen

zu sehen.[17]

Zahlungsmodus	männlich	weiblich
Kreditkarte	33,8%	19,9%
Bankeinzug	26,9%	14,3%
per Nachnahme	38,1%	24,0%
Vorauskasse	2,4%	-
auf Rechnung	31,3%	46,0%
elektron. Zahlungsmittel	1,3%	2,2%
Sonstiges	4,1%	2,9%

[18]

5.4 Zusammenfassung und Ausblick

In dieser Seminararbeit sollte dem Leser ein Einblick in die inhaltsbezogene Abrechnung im Internet ermöglicht worden sein.

Der Wechsel von kostenlosen Diensten hin zu kostenpflichtigen motivierte die Entstehung neuer auf die Gegebenheiten des Internets angepasster Zahlungsmodelle. Wichtig ist hierbei die Unterscheidung von:

- Zahlungsmodellen
- Zahlungssystemen
- Anbietern von Zahlungssystemen

Unter Zahlungsmodellen wird der standardisierte Ablauf eines Zahlungsvorganges im Internet gesehen. Diese Zahlungsmodelle beschreiben den Ablauf der Registrierung, den Erwerb des Zahlungsmittels und den eigentlichen Zahlungsvorgang. Sie können anhand von verschiedenen Eigenschaften klassifiziert werden. Gängige Klassifizierungen sind zum Beispiel:

- nach dem Zeitpunkt der Zahlung
- nach der Bindung an ein Konto

Zahlungssysteme stellen die konkrete Verwirklichung eines Zahlungsmodells dar. Sie sind auf bestimmte Anforderungen an den Bezahlvorgang optimiert und meist nicht universell einsetzbar. Hieraus ergibt sich dann die Nachfrage für Anbieter von Zahlungssystemen, welche wiederum verschiedene Zahlungssysteme kombinieren.

Eine Einführung kostenpflichtiger Internetseiten stellt die Frage nach einer geeigneten Tarifierung. Die Vor- und Nachteile momentaner Tarifmodelle für den Internetzugang lassen sich direkt auf dieses Problem anwenden. Als Beispiel stelle man sich einen Content-Provider vor, welcher eine virtuelle Galerie anbietet. Dem Endverbraucher soll die Möglichkeit gegeben werden sich die Bilder kostenlos anzusehen oder einen kostenpflichtigen Download einzuleiten.

Wenn die angebotenen Bilder jeweils in unterschiedlichen Auflösungen zur Verfügung stehen, stellt sich die Frage, ob ein Nutzer der bereits für ein Bild mit hoher Qualität gezahlt hat und später einen Download desselben Bildes in einer niederen Auflösung initiiert, diesen auch zahlen muss. Intuitiv würde man sagen, er hat bereits den Besitz des Bildes in hoher Qualität erworben also hat er auch ein Anrecht auf dieses Bild in niederer Qualität. Dagegen spricht allerdings die Tatsache, dass ein erneuter Download trotzdem Bandbreite des Content-Providers beansprucht, also für ihn Kosten verursacht.

Das Problem stellt also sich wie folgt dar: Es ist eine Kombination aus einer fairen Tarifierung für den Endverbraucher und einer ökonomischen Abrechnung für den Content-Provider bei seinem ISP zu finden.

Literaturverzeichnis

- [1] "Tarifmodelle im Vergleich", <http://www.fh-wedel.de/cis/archiv/diplomarb/griebel/node5.html#SECTION00212000000000000000>
- [2] "Nicht mehr kostenlos", http://www.outrise.de/kommunikation/internet/strategien/grundlagen/nicht_mehr_kostenlos.html.
- [3] "Heise online", <http://www.heise.de>
- [4] "Weg von der „Kostenlos-Kultur“ im Internet", <http://www.teltarif.de/arch/2001/kw16/s4993.html>
- [5] "ONLINE TODAY", Ausgabe 5/01, EVT 26. April
- [6] "HMD, Heft 224, April 2002", <http://hmd.dpunkt.de/224/01.html>
- [7] "Entscheidungsunterstützung bei der Zahlungssystemwahl im Internet", Von Diplom-Wirtschaftsinformatiker (FH) Martin Reichenbach
- [8] "Bezahlen im Internet", Von Prof. Dr. R. Klein, Fernuniversität-Gesamthochschule-Hagen
- [9] "Internet Open Trading Protocol. Specification Version, No. 0.9.9, OTP, August 1998", <http://www.otp.org/>, aufgerufen am 06.04.1999
- [10] "Phil Janson, Michael Waidner: Electronic Payment Systems. Activity Paper", No. 211ZR018, SEMPER, Mai 1996
- [11] "Semesterarbeit Zahlungsverkehr via Internet in der Schweiz", www.vis.ethz.ch/mnaef/bwl/bericht.pdf
- [12] "oneShell Internet Services GmbH", <http://www.oneshell.de>
- [13] "NZZ online", <http://www.nzz.ch>
- [14] "GBI the contentmachine", <http://www.gbi.de>
- [15] "T-pay", www.t-pay.de
- [16] "Matthias v. Ehrenstein, Bezahlen im WWW, Fernuniversität Hagen", http://gremlin.fernuni-hagen.de/Lehre/Seminare/Hypertext97/kein_rahmen/ehrenstein/#2_11_Anbieter_Akzeptanz

- [17] “Was ist ein EC-System“, Prof. Dr. Gruhn, Universität Dortmund
<http://ls10-www.cs.uni-dortmund.de/LS10/Pages/Lehre-SS99/ec-seminar/traugott-dittmann.ppt>
- [18] “E-Commerce“, Julia Urlinger & Ralph Biermann
http://www.rieffler.net/991222_biermann_urlinger2.ppt.