University of Zurich
Department of Informatics

Burkhard Stiller
Cristian Morariu
Peter Racz
Martin Waldburger
(Eds.)

Internet Economics I

March 2005

TECHNICAL REPORT – No. 2005.05

University of Zurich
Department of Informatics (IFI)
Winterthurerstrasse 190, CH-8057 Zürich, Switzerland

B. Stiller, C. Morariu, P. Racz, M. Waldburger (Eds.):
Technical Report No. 2005.05, March 2005
Communication Systems Research Group
Department of Informatics (IFI)
University of Zurich
Winterthurerstrasse 190, CH-8057 Zurich, Switzerland
URL: http://www.ifi.unizh.ch/csg/

Introduction

The Department of Informatics (IFI) of the University of Zürich, Switzerland started research and teaching in the area of communications. One of the driving topics in applying communications technology is addressing investigations of their use and application under economic constraints and technical optimization measures. Therefore, during the winter term WS 2004/2005 a new instance of the Internet Economic seminar has been prepared and students as well as supervisors worked on this topic.

Even today, Internet Economics are run rarely as a teaching unit. This observation seems to be a little in contrast to the fact that research on Internet Economics has been established as an important area in the center of technology and economics on networked environments. After some careful investigations it can be found that during the last ten years, the underlying communication technology applied for the Internet and the way electronic business transactions are performed on top of the network have changed. Although, a variety of support functionality has been developed for the Internet case, the core functionality of delivering data, bits, and bytes remained unchanged. Nevertheless, changes and updates occur with respect to the use, the application area, and the technology itself. Therefore, another review of a selected number of topics has been undertaken.

Content

This new and first edition of the seminar entitled "Internet Economics I" deals with the use of Internet technology and additional ways to support and do business. Starting the talks, a view onto "Commercialized Real-time Transport of Data across the Internet" was presented. This first talk addressed the details on real-time transport and outlined some selected protocols, streaming media, and their discussion in a market situation. Talk two presented a view onto "IP Convergent Networks", which argues on convergence influences and sees the Voice-over-IP application as the driving force for a successful convergence. The third talk addresses "Trust and Identity", which are described in terms of applications and their requirements, social aspects, technology basics, and legal issues. "Technology- and Layer-independent Authentication Schemes for the Internet" as talk four discuss those options, which provide protocols for an authentication independent of the underlying technology utilized. In particular, the proposal of PANA is described and investigated under certain assumptions. The talk number five summarizes frequent types of attacks that are known today, which may result in damage in a company's network infrastructure. A closer look into Distributed DoS attacks is presented.

The sixth talk "Das Individuum als Operator" outlines the situation of today's operators and argues on historic developments. It provides additionally, viable models for operators to survive market trends – focused on the Wireless LAN segment – and technology changes, a.o. driven by disruptive technologies. Talk number seven addresses "New Opportunities for B2B and B2C Services Offered by New Mobile Communication Technologies" overviews current wireless technologies, describes the current market situation with respect to content, applications, pricing, and payments, and forecasts into the 4th generation. "Gruppenstrategien für Online-Business am Beispiel Mobilfunk" within the eights talk presents alliance forming reasons and their economic as well as legal aspects. Applied to the telecommunications market three phases and their alliance behavior are discussed. The talk number nine on "Overview and Analysis of Content Provider Business" structures the term of a business model with respect to the content term and discusses Rappa's model in greater detail due to its promising nature. Case studies outline the usability of this mode. Talk ten continues with "Verrechnungsmethoden für Inhalte aus dem Internet" and outlines existing payment methods, classifies those, and evaluates their chances to be introduced into today's content market. Finally, talk number 11 on "Migration to IPv6" provides a technology insight into the key differences of IPv4 and IPv6 and discusses the migration issue.

Seminar Operation

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

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

Local IFI support for preparing talks, reports, and their preparation by students had been granted Cristian Morariu, Peter Racz, Martin Waldburger, and Burkhard Stiller. In particular, many thanks are addressed to Peter Racz for his strong commitment on getting this technical report ready and quickly published. A larger number of pre-presentation discussions have provided valuable insights in the emerging and moving field of Internet Economics, both for all groups of students and supervisors. Many thanks to all people contributing to the success of this event, which has happened in a lively group of highly motivated and technically qualified students and people.

Zürich, February 2004

Contents

Kapitel 1

Commercialized Real-time Transport of Data across the Internet

Emanuel Giger, Martin Morger, Sébastien Donzel

"Das Medium ist die Botschaft" (McLuhan 1992 [25]). Der stetige Wandel der Technik (Medium) hat Auswirkungen auf die gesellschaftlichen und ökonomischen Muster und Formen (Inhalt). Die Botschaft eines Mediums oder einer Technik ist die Veränderung des Massstabs, Tempos oder Schemas, die es den Menschen bringt. Die Umstellung von analog verarbeiteter und verbreiteter Information hin zur vollumfänglichen Digitalisierung dieser ermöglichen neue Geschäftsmodelle. Real-Data Transfer basierend auf dem RTP und RTSP Protokoll ermöglicht eines dieser neuen digitalen Geschäftsmodellen, welchen wir in dieser arbeit auf den Grund gehen möchten. Wir zeigen die Funktionsweisen der beiden Protokolle im Zusammenhang mit dem Audio-/Video Streaming über das Internet auf. Des Weiteren wird gezeigt, inwiefern die Protokolle in neuen Geschäftsmodellen implementiert wurden, und basierend auf dieser Grundlage inwieweit diese Modelle einen kommerziellen Nutzen mit sich bringen. Unserer Analyse nach steckt die gesamte Kommerzialisierung von Real-Time Data Transfer über das Internet noch in den Kinderschuhen und wir räumen ihr, obwohl des enormen Potentials, eher wenig Chancen ein, solange in Sachen Qualität, Preis und Urherberrechtslage keine Verbesserungen gemacht werden.

Inhaltsverzeichnis

1.1 Einleitung

Die Verbreitung von multimedialen Daten über das Internet erfreut sich grosser Beliebtheit. Dabei sind kurze Wartezeiten und eine gute Qualität der Datenpräsentation entscheidende Faktoren. Den Inhalt komplett aus dem Internet zu laden, bevor mit der Wiedergabe begonnen werden kann, ist daher nicht gerade eine angemessene Vorgehensweise. Wünschenswert wäre ein Verfahren, dass unmittelbar mit dem Abspielen der Daten beginnt und sogar die Möglichkeit bietet, wie bei einem CD- oder DVD-Player, beliebig durch den Inhalt zu navigieren.

Ein mögliches Lösungskonzept bietet die Streaming-Technologie. Sie erlaubt Audio- und Videodaten, aus dem Internet in Echtzeit zu übertragen. Es ist nicht mehr notwendig, die ganze Datei aus dem Internet vollständig auf den Rechner zu laden, bevor mit der Inhaltswiedergabe begonnen werden kann. Nur eine kleine Menge an Daten wird benötigt, um mit der Wiedergabe zu beginnen. Die restlichen Daten werden während der Übertragung laufend nachgeladen. Dadurch werden auch Live-Übertragungen möglich.

Eine solche Art der Datenübertragung erfordert gewisse technische Voraussetzungen. Einerseits müssen grosse Datenvolumen auf eine Menge reduziert werden, die an eine Internetübertragung angemessen ist.[1] Dies verlangt nach einer ausgefeilten Technik von Komprimieralgorithmen. Vor allem der MPEG-Standard [1] erlaubte erstmals eine internettaugliche Datenkomprimierung. Bei der Hardware andererseits sind spezielle Streaming-Server und eine gewisse Mindestbandbreite notwendig, um selbst komprimierte Daten in einer für den Benutzer akzeptablen Qualität zu übertragen. Durch die zunehmende Verbreitung von schnellen Breitbandanschlüssen bei Privathaushalten wurde dieses Problem in den letzten Jahren entschärft.

Die Komprimierung ist jedoch nur eine notwendige Bedingung für eine Übertragung von Daten im Internet, die sich auf Grund der Kapazität heutiger Netze aufdrängt. Damit die Daten wirklich übertragen werden können, sind Protokolle notwendig, die einen Standard für eine Übertragung darstellen. Im Kontext von Streaming war die Standardisierung von Protokollen speziell für Streaming und Echtzeitdatenübertragung ein logischer und notwendiger Schritt. Eine einheitliche Weiterentwicklung in diesem Gebiet wäre sonst kaum möglich.

Die vorliegende Arbeit behandelt die Thematik des kommerziellen Realtime-Streaming an Hand von zwei Betrachtungsansätzen: Im Kapitel 2 wird ein „Bottom-Up“-Ansatz gewählt, der die Technik erläutert, die Realtime-Streaming ermöglicht. Kapitel 3 hingegen geht mit einem „Top-Down “-Ansatz auf die kommerziellen Aspekte ein. Ziel ist es zu zeigen, ob beide Ansätze sich irgendwo treffen, und somit Technik und kommerzielle Anwendung in einer Beziehung zueinander stehen, in der sie sich ein- oder gegenseitig beeinflussen.

[1] Die Grösse der Daten spielt natürlich bei allen Arten der Übertragungen eine Rolle. Hier handelt es sich jedoch um Audio/Video-Daten, die in der Regel besonders gross sind.

1.2 Technologie

Zu Beginn einer Streaming Anwendung stehen die Nutzdaten. Da die unbehandelten Daten in der Regel zu gross für eine Übertragung im Internet sind, müssen sie erst auf eine Grösse reduziert werden, die für eine Übertragung praktikabel ist. Die Komprimierung geschieht durch Codecs („Compressor / Decompressor “), wie das im nachfolgenden Kapitel erläuterte MPEG [1]. Erst nach der Komprimierung werden die Daten auf einen Media- oder Webserver abgelegt, von welchem aus sie über das Internet abrufbar sind. Protokolle regeln bei Abruf die Übertragung der Daten zwischen Server und Client. Bevor die Daten auf der Clientseite präsentiert werden können, müssen sie zuerst mit dem entsprechenden Codec dekomprimiert werden. Nach der Dekomprimierung ist ein Media Player des Clients in der Lage die Daten abzuspielen. `Abbildung 1.1` zeigt den typischen Ablauf einer Streaming- Anwendung und den Einsatz von Codecs.

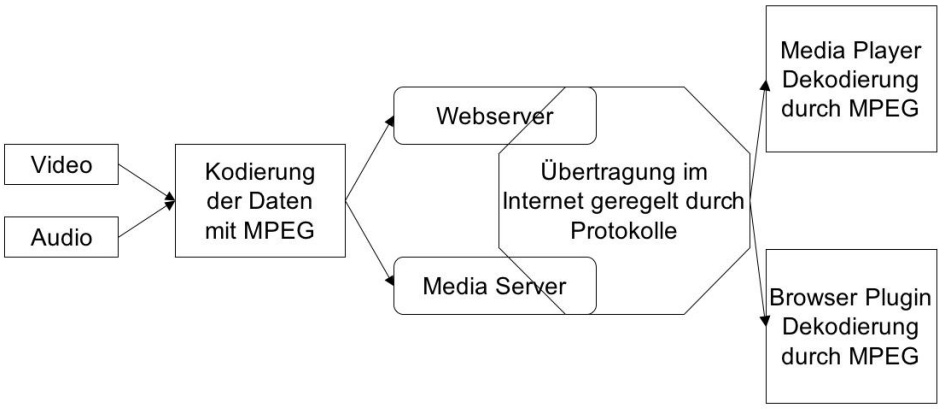

Abbildung 1.1: Ablauf einer Streaming-Anwendung

Um die zeitsensitiven Daten von Streaming Anwendungen zu übertragen, sind während des Transports zusätzliche Massnahmen notwendig. Da weder TCP (Transmission Control Protocol) [36] noch UPD (User Datagram Protocol) [39] die Übertragung von Echtzeitdaten in einer angemessenen Art und Weise unterstützen, musste jede Anwendung, die Daten in Echtzeit übertragen wollte, ein eigenes Protokoll implementieren. Die einzelnen Protokolle hatten alle dasselbe Ziel, waren untereinander jedoch nicht kompatibel. Daher wurden vier Standards entwickelt, welche die zukünftige Entwicklung von Streaming-Anwendungen vereinheitlichen:

- Realtime Transport Protocol (RTP) [28]: RTP übernimmt den Datentransport. Mit dem Realtime Transport Control Protocol (RTCP) besitzt RTP neben dem Datenteil einen Kontrollteil. RTP/RTCP ist in RFC 1889 definiert und wird im Folgenden genauer betrachtet.

- Realtime Streaming Protocol (RTSP) [29]: RTSP steuert den Datenstrom und ermöglicht einen effizienten Datentransport über IP-basierte Netzwerke. Funktionsweise und Syntax gleichen HTTP. RTSP ist in RFC 2326 definiert und nachfolgend ebenfalls detailliert erläutert.
- Resource Reservation Protocol (RSVP) [30]: RSVP reserviert die notwendigen Netzressourcen für eine Datenübertragung. Der Empfänger erhält die Möglichkeit, eine bestimmte Menge an Ressourcen für den Datenstrom zu reservieren. Die Reservation wird von allen Routern bis hin zum Empfänger sichergestellt - vorausgesetzt die Router implementieren RSVP. RSVP wurde von der RSVP Working Group in RFC 2205 definert.
- Synchronisation Multimedia Integration Language (SMIL) [35]: SMIL ist eine auf XML basierende Sprache, die der Formatierung und Beschreibung des Medienstreams dient. Sie ermöglicht eine Synchronisation verschiedener Medien innerhalb einer Anwendung. SMIL wird vom World Wide Web Consortium (W3C) [41] betreut

Für den Transport von Streaming Daten über das Internet können zwei verschiedene Protokolle der Transportschicht des TCP/IP Protokollstapels verwendet werden, welche beide auf dem Internet Protokoll [15] aufsetzen. Die Adresse eines Endpunktes besteht bei TCP und UDP aus der Kombination von IP-Adresse und einer Portnummer.

TCP

Das Transmission Control Protocol [36] ist ein zuverlässiges, verbindungsorientiertes Transportprotokoll. Der Header besteht aus insgesamt 18 Feldern. TCP stellt einen virtuellen Kanal zwischen Sender und Empfänger her, so dass jedes Paket die gleiche Route befolgt. Die Integrität der Daten wird mittels einer Prüfsumme, die Reihenfolge der Pakete mit einer Sequenznummer sichergestellt. Die zuverlässige Zustellung der Pakete wird garantiert, in dem für jedes Paket vom Empfänger eine Empfangsbestätigung versendet wird. Der Sender wiederholt das Senden des Paketes so lange, bis er eine Empfangsbestätigung erhält hat bzw. ein Timeout auftritt.

UDP

Das User Datagram Protocol [39] ist ein minimales, verbindungsloses Transportprotkoll. Die Einhaltung der Reihenfolge und der verlustfreie Empfang der Pakete sowie ein konstanter Jitter können nicht gewährleistet werden. Ein verlorenes Paket wird vom Sender anders als bei TCP nicht nochmals übertragen. Bei Verzögerungen auf Netzabschnitten zwischen Sender und Empfänger wird ein UDP-Paket über eine alternative Strecke weitergeleitet, da keine festgelegte Route zu befolgen ist.

Diesen Nachteilen steht der für die Übertragung von Multimediadaten wichtige Vorteil gegenüber, dass die Daten einfacher verarbeitet und damit schneller übertragen werden können als mit einem zuverlässigen Protokoll. Daher wird für den Transport von Multimediastreams häufig UDP statt TCP eingesetzt. Die Protokolle der Anwendungsschicht

stellend dabei die Einhaltung der Reihenfolge der Pakete sicher und ermöglichen durch eine Pufferung der Daten beim Empfänger eine gewisse Glättung von Jitter.

1.2.1 MPEG

Ein Spielfilm von 90 Minuten Länge mit hoher Auflösung und vielen Farben hat eine Grösse von ca. 120 GByte. Mit diesem Datenvolumen lässt sich der Film weder auf einer herkömmlichen DVD noch auf einer CD speichern. An eine Übertragung im Internet ist gar nicht zu denken. Gesucht sind daher Wege, welche die Datengrösse eines Spielfilms um ein Vielfaches verkleinern, ohne die Qualität beim Zuschauer wahrnehmbar zu reduzieren. Erst Komprimierverfahren ermöglichen eine Nutzung von Film in Verbindung mit dem Internet. Mit “MPEG“ wird einer der bekanntesten Standards in diesem Abschnitt erläutert.

MPEG steht für “Motion Picture Experts Group“ und bezeichnet eine Expertengruppe, die sich mit Standards zur komprimierten Speicherung von Audio- und Videodaten befasst [37]. Im Alltagsgebrauch wird MPEG auch als Name für das Verfahren selbst benutzt.

MPEG ist ein asymmetrisches Kodierungsverfahren, was bedeutet, dass die Kodierung aufwendiger als die Dekodierung ist. Von MPEG existieren vier verschiedene Standards für unterschiedliche Anwendungen. MPEG wird hier am Beispiel von MPEG 1 betrachtet. MPEG 1 wurde entwickelt, um Filme bis auf eine Rate von 1.5 MBit/s zu komprimieren und wird beispielsweise für Video-CDs verwendet. Das Ergebnis ist von eher bescheidener Qualität. Der MPEG 2 Standard, der für DVD benutzt wird, besitzt mit einer Datenrate zwischen 15MBit/s und 50 MBit/s bereits eine bedeutend bessere Qualität.

MPEG 1 ist eine Gruppe von Verfahren. Der zentrale Punkt von MPEG ist, dass nicht versucht wird, die einzelnen Bilder möglichst stark zu komprimieren, sondern die Differenzen von aufeinander folgenden Bildern. MPEG bedient sich damit der Eigenschaft, dass sich aufeinander folgende Bilder in einer Bildsequenz bis zu einem gewissen Grad gleichen.

Motion Compensation

Ein zentraler Bestandteil von MPEG ist Motion Compensation. Dabei wird das aktuelle Bild in Makroblöcke unterteilt. Das nächste Bild in der Folge wird nach diesen Blöcken durchsucht. Wenn sich nun zwischen den Bildern wenig ändert, können die Blöcke des aktuellen Bildes im nachfolgenden Bild wieder gefunden werden. Dabei wird nicht nur die gleiche Position im nachfolgenden Bild auf Übereinstimmung untersucht, sondern ein bestimmter Ausschnitt, um die Bewegung der Blöcke in einer Bildfolge zu kompensieren. Daher der Name Motion Compensation. Der räumliche Unterschied wird durch einen Bewegungsvektor beschrieben.

Blöcke in Bildfolgen können sich nicht nur räumlich unterscheiden, sondern auch inhaltlich. Die Unterschiede in Helligkeit und Farbe werden durch ein Differenzbild festgehalten. Der einfachste Algorithmus wäre, einen gegebenen Block des aktuellen Bildes im nachfolgenden Bild innerhalb eines bestimmten Bewegungsfeldes zu suchen. Als Suchkriterium kann die

Höhe der Übereinstimmung in Helligkeit und Farbe benutzt werden. Da nur die Differenz relevant ist, muss bei hundertprozentiger Übereinstimmung nichts gespeichert werden. Was noch als gefunden gilt, hängt von der jeweiligen MPEG-Implementierung ab. Kann keine Übereinstimmung innerhalb des Toleranzbereiches gefunden werden, ändert sich zu viel, und das Bild muss vollständig gespeichert werden. In diesem Fall wird das Bild wie ein I-Frame (vgl. `Kap. 1.2.1`) gespeichert. Die Implementierungen können sich daher je nach Bedarf in Qualität und Geschwindigkeit unterscheiden. Sie sind jedoch alle kompatibel, da MPEG nur das Dateiformat und nicht die Kompressionsalgorithmen vorschreibt [37].

MPEG Bildtypen

MPEG benutzt vier verschiedene Bildtypen, um eine Bildsequenz zu kodieren [37]:

- I-Frames (Intra-coded Pictures): I-Frames sind unabhängig von anderen Bildern der Sequenz. I-Frames werden wie einzelne Bilder betrachtet und im JPEG-Format gespeichert. Es findet keine Bewegungskompensation statt. Die Kompressionsrate ist mit 1:7 im Vergleich zu P-Frames und B-Frames eher gering. Beim Rechnen mit Differenzen ist mindestens am Anfang ein absoluter Wert notwendig, um mit Hilfe der Differenz den aktuellen Wert zu erhalten. I-Frames dienen daher als Ankerpunkte, an denen ein Neueinstieg möglich ist z. B. bei einer Live-Übertragung oder bei einem Neubeginn der Dekodierung, falls vorher fehlerhafte Rahmen übertragen wurden.
- P-Frames (Predictive-coded Pictures): Sie benutzen Motion Compensation und sind daher jeweils von zeitlich vorausgehenden I- oder P-Frames abhängig. Die Kompressionsrate beträgt 1:20. Wenn eine genügende Übereinstimmung von Makroblöcken aufeinander folgender Bilder betreffend Inhalt gefunden wurde, so enthält der Makroblock nur die Differenzen zum Vorgänger. Zusätzlich wird in einem P-Frame die räumliche Differenz als Bewegungsvektor kodiert. Der Differenzblock wird der diskreten Kosinustransformation (DKT), der Quantisierung, der Längenlaufkodierung und der Huffman-Kodierung unterzogen und in dieser Art weiter komprimiert, indem unter anderem auf nicht wahrnehmungsrelevante Bildinformationen verzichtet wird. Bei der DKT entstehen Rundungsfehler, die MPEG 1 zu einem verlustbehafteten Verfahren machen.
- B-Frames (Bidirectional predictive-coded Pictures): B-Frames beziehen sich sowohl auf ein vorhergehendes als auch auf ein nachfolgendes Bild. Sie bieten mit 1:50 den grössten Kompressionsgrad.
- D-Frames (DC-coded Pictures): D-Frames werden bei schnellem Navigieren im Video eingesetzt. Sie arbeiten mit niedriger Auflösung, da eine Echtzeitdekodierung in normaler Qualität im schnellen Durchlauf kaum möglich ist.

MPEG schreibt keine bestimmte Reihenfolge vor, in welcher die verschiedenen Bildtypen übertragen werden müssen. In der Regel stimmt die Bildsequenz im MPEG-Datenstrom nicht mit der Abspielreihenfolge überein. Weiterführende Informationen zu MPEG und den verwendeten Kompressionsalgorithmen findet man der offiziellen MPEG-Homepage MPEG.

1.2.2 Realtime Transport Protocol (RTP) [28]

RTP wird in der Regel über UDP [39] ausgeführt, um dessen Multiplex-Fähigkeit und die Prüfsummen im UDP-Header zu nutzen. Der RTP-Header und die RTP-Nutzdaten werden als RTP-Paket in die Nutzdaten eines UDP-Paketes verpackt.

So gesehen ist die Stellung von RTP im Protokollstapel nicht eindeutig: Einerseits arbeitet RTP im Benutzerbereich auf UDP basierend und ist mit der Anwendung verknüpft. Andererseits stellt RTP einen allgemeinen, unabhängigen Standard zur Übertragung von Echtzeitdaten dar und verhält sich demnach wie ein Transportprotokoll. `Abbildung 1.2` zeigt die etwas spezielle "eingeschobene" Stellung von RTP innerhalb des 4-Ebenen-Stappels, der typischerweise heute im Internet verwendet wird. RTP unterstützt sowohl Unicast wie auch Multicast.

Mit RTCP (Realtime Transport Control Protocol) besitzt RTP ein "kleines Schwesterprotokoll", das ebenfalls in RFC 1889 definiert ist. RTCP unterstützt RTP in dessen Aufgabe, indem es permanente Inforationen über die verfügbaren Dienstqualitäten und über die Teilnehmer der Sitzung bereitstellt (vgl. `Kap. 1.2.2 und Kap. 1.2.3`).

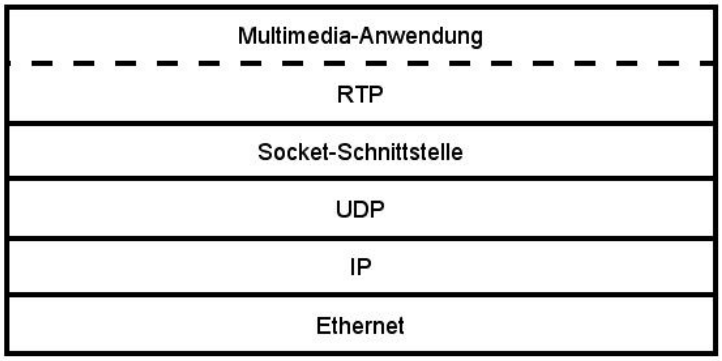

Abbildung 1.2: Stellung von RTP/RTCP im Protokollstapel

Aufgaben von RTP und RTCP

Während der Übertragung müssen die Echtzeitdaten in Pakete aufgeteilt und getrennt versendet werden. Dies impliziert gewisse Aufgaben, die für den Transport von Echtzeitdaten erfüllt werden müssen:

- Die Pakete müssen beim Empfänger in der ursprünglichen Reihenfolge sortiert vorliegen.
- Verschiedene Medienströme müssen eventuell synchronisiert werden.
- Anpassung an variable Dienstqualitäten (QoS), damit ein kontinuierlicher Datenstrom möglich ist.

- Es werden Informationen (Meta-Daten) über die Teilnehmer der Übertragung benötigt.

RTP und RTCP stellen Funktionen zur Verfügungen mit welchen die genannten Aufgaben erfüllt werden können:

- Kennzeichnung der übertragenen Nutzdaten und ihrer Quelle
- Sequenznummern
- Zeitstempel
- Kontrolle der vorhandenen Dienstqualitäten
- Übertragung von Informationen über die verschiedenen Teilnehme

Die letzten beiden Funktionen werden von RTCP übernommen.

RTP ist weder für die Einhaltung von Dienstqualitäten zuständig, noch ist garantiert, dass die Pakete beim Empfänger synchron, in der richtigen Reihenfolge oder überhaupt ankommen. Dies muss durch andere Protokolle oder die Anwendung selbst mit Hilfe der Möglichkeiten, die RTP und RTCP bieten, sichergestellt werden.

Protokoll-Architektur

In der Regel werden Protokolle durch erweiterbare und optionale Felder im Header sehr allgemein gehalten. Dies ermöglicht eine gewisse Flexibilität in Bezug auf zukünftige Anpassungen und Erweiterungen des Protokolls. Dadurch wird aber ein aufwändiges Parsen des Headers notwendig. Für zeitsensitive Daten ist dies ungeeignet. Daher werden die spezifischen Daten der Anwendung nicht im Header selbst, sondern in Profilen für die jeweiligen Anwendungen definiert. RTP selbst behandelt nur Aspekte, die alle Anwendungen betreffen. Das derzeit einzig weit verbreitete Profil ist in RFC 1980 [32] definiert und dient der Audio- und Videoübertragung z.B. in Konferenzen. Innerhalb eines Profils sind die zulässigen Nutzdatentypen bzw. deren Kodierungen definiert. Da die Informationen über Nutzdaten mit jedem Paket im Header gesendet werden, ist es möglich, die Kodierung dynamisch zu wechseln, um beispielsweise auf schwankende Bandbreiten zu reagieren. Diese Systemarchitektur setzt jedoch zusätzliche Spezifikationen für Profile und Nutzdatentypen voraus. `Abbildung 1.3` zeigt die Protokoll-Architektur, die RTP/RTCP zu Grunde liegt.

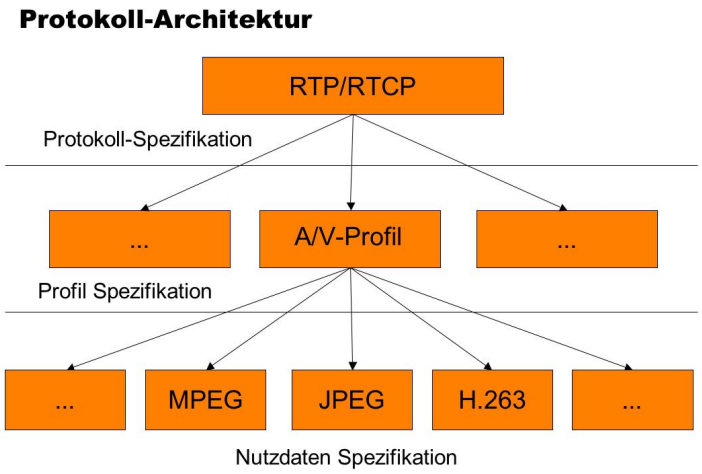

Abbildung 1.3: Protokoll-Architektur bei RTP/RTCP

RTP-Header

Abbildung 1.4: RTP-Header

Erklärung der in `Abbildung 1.4` dargestellten Feldern:

- Version (V): Bezeichnet die Version von RTP. Aktuell ist Version 2.
- Extension (X): Falls gesetzt, ist der Erweiterungsheader vorhanden.
- Contributing Source Count (CC): Anzahl der vorhandenen beteiligten Quellen. Beteiligte Quellen kommen beispielsweise bei Verwendung eines Mixers vor.[2]

Für die Übertragung von Echtzeitdaten sind folgende Headerfelder von besonderer Bedeutung:

- Nutzdatentyp: Beschreibt die übertragenen Daten.

[2]Ein Mixer kombiniert die Datenströme mehrerer Teilnehmer zu einem einzelnen Datenstrom.

- Folgenummer: Die Folgenummer ist ein simpler Zähler, der mit jedem Paket erhöht wird. Dies erlaubt dem Empfänger, den Verlust von Paketen zu erkennen. Bei verlorenen Paketen findet keine Neuübertragung statt, da dies bei Echtzeitdaten zu unzumutbaren Verzögerungen führen würde. Es wird stattdessen versucht, mittels Interpolation oder anderen Verfahren den fehlenden Wert zu ergänzen.
- Zeitstempel: Bezeichnet den Erstellungszeitpunkt der Nutzdatenbytes. Der Zeitstempel wird benutzt, um mehrere Medienströme z.B. Audio- und Videodaten einer Videoübertragung zu synchronisieren.
- Synchronisationsquelle: Identifiziert die Quelle der Übertragung.

1.2.3 Realtime Transport Control Protocol (RTCP) [28]

Aufgaben von RTCP

RTCP unterstützt RTP bei der Übertragung von Echtzeitdaten. Das darunter liegende Protokoll muss Multiplexing unterstützen, da RTP- und RTCP-Daten in der gleichen Sitzung übertragen werden.

Eine der Hauptaufgabe von RTCP ist das Senden von Informationen über die Dienstqualitäten. RTCP überträgt dazu periodisch Sender-Reports (SR) oder Receiver-Reports (RR).Dies ermöglicht den teilnehmenden Quellen, ihren Datenstrom der Netzbelastung anzupassen. So kann beispielsweise die Übertragungsrate bei tiefer Dienstqualität reduziert werden. Das Senden solcher Überwachungsdaten ist auch an Drittparteien möglich, zum Beispiel an einen externen Überwachungsprozess.

Die zweite wichtige Aufgabe von RTCP ist das Ermöglichen einer permanenten Identifizierung der Teilnehmer. Dazu überträgt RTCP ebenfalls periodisch ein Source-Description-Paket, das eine eindeutige Quellidentifikation sowie eine Reihe anderer Informationen (Telefonnummer, E-Mail, Standort etc.) enthält.

Multiplexing verschiedener Medienströmen

RTP/RTCP selbst bietet keinen Multiplexmechanismus an. Ströme verschiedener Medien werden über unterschiedliche RTP-Sitzungen übertragen. Die Zusammenführung unterschiedlicher Medien der selben Quelle, z.B. Audio- und Videostrom bei einer Videokonferenz, erfolgt über IP-Adresse, Port-Paar und die Identifikation im RTCP-Paket durch die Anwendung. Die Synchronisation der einzelnen Medienströme untereinander erfolgt über den Zeitstempel im RTP-Header.

Es gibt einige Gründe, weshalb RTP/RTCP keinen eigenen Multiplexmechanismus anbietet:

- Eine aufwändigere Implementierung des Protokolls bei Bereitstellung eines solchen Mechanismus wäre die Folge.

- Getrennte Sitzungen können unterschiedliche Netzpfade benutzen.
- Bei geringer Bandbreite kann ein Datenstrom fokussiert werden, um bei Engpässen beispielsweise nur noch den Audiostrom einer Videokonferenz zu übertragen.

`Abbildung 1.5` zeigt Multiplexing über RTP und RTCP am Beispiel einer Audio/Video-Konferenz.

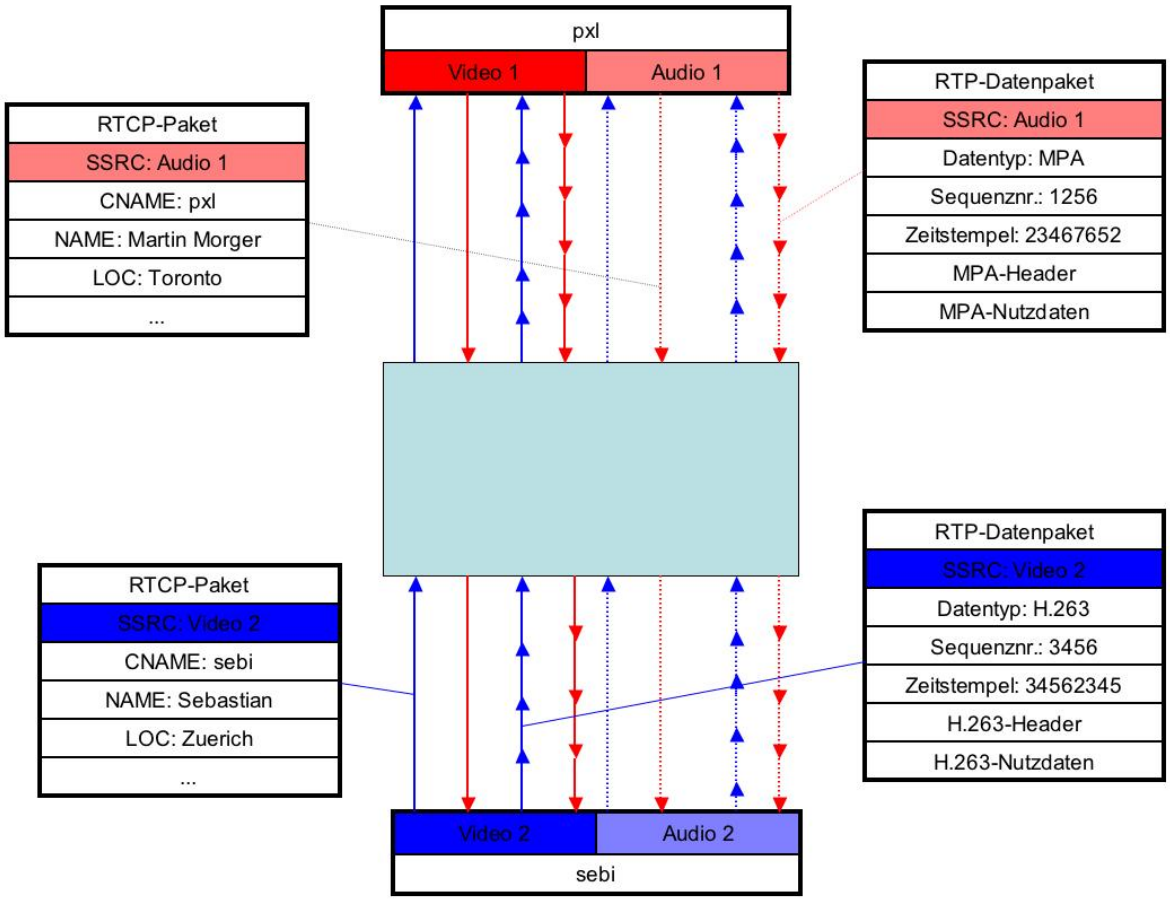

Abbildung 1.5: Multiplexing verschiedener Medienströme

Die Audio- und Videodaten werden in getrennten Sitzungen, die jeweils über ein eigenes RTP/RTCP-Paar verfügen, übertragen. Die Verknüpfung der verschiedenen Sitzungen erfolgt über die Teilnehmerkennung, die in den RTCP-Paketen übertragen wird sowie die IP-Adresse und das Port-Paar. Die Synchronisation der beiden Medienströme geschieht mit Hilfe des Zeitstempels im RTP-Paket.

1.2.4 Real Time Streaming Protocol (RTSP) [29]

Das Real Time Streaming Protocol (RTSP) wurde gemeinsam von RealNetworks, Netscape Communications und der Columbia University im Rahmen der IETF Music Group

[14] entwickelt. Im April 1998 wurde es von der Internet Engineering Task Force (IETF) als Proposed Standard veröffentlicht und als RFC 2326 [29] dokumentiert.

Bei RTSP handelt es sich um ein Protokoll der Anwendungsschicht des TCP/IP Protokollstacks, welches zum Aufbau und zur Kontrolle von einem oder mehreren zeitsynchronen Multimediaströmen (Streams), typischerweise Audio- und Videoströme verwendet wird. Die übertragenen Ströme können sowohl Echtzeit Livedaten als auch gespeicherte Aufzeichnungen enthalten. Die Übertragung der Medienströme erfolgt dabei nicht durch RTSP selbst, sondern mit Hilfe des Realtime Transport Protokoll [28] (siehe auch Kapitel 2.2).

RTSP kann somit vereinfacht als eine Art Netzwerk-Fernbedienung für Multimedia-Server bezeichnet werden. Die Menge von einem oder mehreren Strömen, welche dem Benutzer als ein zusammenhängendes Multimediaobjekt präsentiert werden (z.B. ein Video- und mehrere Audioströme), wird als Präsentation bezeichnet. Im Falle eines Videos mit Stereoton besteht eine solche Präsentation beispielsweise aus einem Video- und zwei Audioströmen. Die Präsentation wird durch eine sog. Präsentations-Beschreibung definiert. Welches Format für diese Beschreibung verwendet wird, ist nicht Bestandteil der RTSP-Spezifikation, als verbreiteter Standard für solche Präsentations-Beschreibungen wird jedoch häufig das Session Description Protokoll [34] verwendet.

Bei RTSP wird keine virtuelle Verbindung zwischen Client und Server aufgebaut wie bei TCP. Stattdessen eröffnet der Server für jeden Client eine sog. Session, welche durch eine eindeutige Session-ID identifiziert wird. Während einer Session kann ein Client beliebige zuverlässige Transportverbindungen (z.B. TCP) mit dem Server aufbauen und wieder schliessen oder verbindungslose Protokolle (wie UDP) verwenden, um durch diese Verbindungen RTSP-Befehle, wie 'stop' oder 'play' abzusetzen.

Abgrenzung Codec / RTSP

RTSP kontrolliert die Übertragung der Medienströme, die aus Paketen bestehen, welche wiederum die Container für die eigentlichen Mediennutzdaten sind. Beispiele möglicher Containerformate sind RTP (Realtime Transport Protocol), ASF (Advanced Streaming Format, Microsoft), RealMedia (Real Networks) oder Quicktime (Apple). Die Nutzdaten werden mit Hilfe eines Codecs (Codierer/Decodierer) codiert, dabei werden die Nutzdaten komprimiert und mit Fehlerkorrekturdaten ergänzt. Beim Empfänger werden die Daten wieder entsprechend decodiert. Eine Reihe von verschiedenen Codecs stehen für Multimediadaten zur Verfügung, so z.B. MPEG-Video, DivX, RealVideo und Windows Media Video für Videodaten sowie MP3, RealAudio und Windows Media Audio für Audiodaten.

Syntax

Die Syntax RTSP wurde in Anlehnung an das Hypertext Transport Protokoll [16] entwickelt, so dass Erweiterungen von HTTP auch zu RTSP hinzugefügt werden können. Allerdings sind folgende Unterschiede zwischen RTSP und HTTP zu beachten:

- RTSP verwendet die Protokoll-Kennung RTSP/1.0
- RTSP-Server sind anders als HTTP-Server nicht zustandslos sondern zustandsbehaftet. Der Zustand jeder Sitzung wird mit Hilfe des Session Identifiers gespeichert.
- Im Gegensatz zu HTTP, wo nur der Client Anfragen an den Server senden kann, ist es bei RTSP auch dem Server möglich, Anfragen an den Client zu senden.
- Ein RTSP-Request verwendet immer absolute URLs, im Gegensatz zu HTTP, welches auch relative Pfade unterstützt.
- RTSP definiert ausserdem einige zusätzliche Methoden, die HTTP nicht unterstützt, und verwendet einen anderen Zeichensatz (ISO 10646) als HTTP (ISO 8859-1) Das RTSP-Protokoll bietet im Wesentlichen drei Operationen:
- Empfang von Medienströmen von einem Server
- Hinzufügen eines Medienservers zu einer bestehenden Konferenz (zur Wiedergabe von zusätzlichen Medien oder zur Aufnahme der Präsentation oder einzelnen Medien davon)
- Hinzufügen von zusätzlichen Medien zu einer Präsentation

Eigenschaften

Im Folgenden eine Auswahl der charakteristischen Eigenschaften von RTSP (für eine komplette Liste siehe RFC 2326, Abschnitt 1.4 [29])

Durch die Erweiterbarkeit lassen sich bei Bedarf neue Methoden und Parameter einfach zum bestehenden Protokoll hinzufügen. Als Sicherheitsmechanismen werden die beiden HTTP-Sicherheitstechniken Basic und Digest-Authentisierung unterstützt, zusätzlich sind auch weitere Sicherheitsfunktionen der Transport- und der Netzwerkschicht anwendbar.

Die Unabhängigkeit vom Transportprotokoll ermöglicht die Verwendung sowohl von unzuverlässigen Datagrammprotokollen wie UDP als auch von zuverlässigen Protokollen wie TCP für die Übertragung der RTSP-Befehle und der Nutzdaten. Befehle und Nutzdaten können dabei unterschiedliche Protokolle verwenden, so dass die Befehle meist über TCP und die Nutzdaten über UDP übertragen werden.

RTSP unterstützt Multiserver-Fähigkeit, so dass jeder Medienstrom einer Präsentation falls nötig von einem anderen Server bezogen werden kann. Der Client stellt dabei die benötigten Kontrollverbindungen mit den Servern her. Dadurch ist ein gewisser Lastausgleich möglich, indem beispielsweise die Audioströme auf einem anderen Server abgelegt werden als der Videostrom einer Präsentation.

Schliesslich bietet RTSP neben der Wiedergabe der Medien auf dem Client auch die Möglichkeit einer Aufnahmefunktion auf dem Server.

Präsentations-Beschreibung

Jede Präsentation und jeder Medienstrom kann durch eine eigene RTSP-URL identifiziert werden. Die Eigenschaften der Präsentation sowie der einzelnen Medien, aus denen die Präsentation besteht, werden mit SDP in einer Präsentations-Beschreibungs-Datei definiert.

Diese SDP-Datei wird vom Client typischerweise über HTTP bezogen und enthält Angaben zur Präsentation und den darin enthaltenen Medien wie eine allgemeine Beschreibung, verwendete Codierung und (natürliche) Sprache des Inhaltes.

In der SDP-Datei wird jeder durch RTSP steuerbare Medienstrom durch eine eigene RTSP-URL identifiziert, welche den Pfad zum Medienserver und den Namen des Stromes auf diesem Server angibt. Für eine vollständige Liste der Attribute von SDP, siehe RFC 2327 [34].

Nachrichten

Die Teilnehmer einer RTSP-Sitzung kommunizieren untereinander über RTSP-Nachrichten, wobei zwischen Anfragen (Requests) und Antworten (Responses) unterschieden wird. Dabei ist es bei RTSP auch dem Server möglich, eine Anfrage an den Client zu schicken. Einige Methoden sind jedoch nur auf dem Server bzw. nur auf dem Client aufrufbar.

RTSP ist ein textbasiertes Protokoll. Das Format der Nachrichten entspricht dem Format der in RFC 2616 [16], Abschnitt 4 beschriebenen HTTP-Nachrichten. Sie bestehen analog zu HTTP aus einer Startzeile, einer oder mehreren Kopfzeilen (Header) die durch eine Leerzeile abgeschlossen werden und einem optionalen Nachrichtenkörper (Body).

Die Trennung der einzelnen Nachrichten erfolgt durch die Kombination der ASCII-Zeichen CR und LF. Enthält die Nachricht einen Nachrichtenkörper, so wird dessen Länge durch eine Content-Length Kopfzeile angegeben. `Die Abbildungen 1.6 und 1.7` zeigen einen RTSP-Request des Clients und die darauf folgende Response des Servers. Beide enthalten nur eine Startzeile sowie Kopfzeilen, der Nachrichtenkörper ist jedoch leer.

Methoden

Die Befehle an den Server oder an den Client werden als sog. Methoden übermittelt, die von RTPS unterstützten Methoden sind in `Abbildung 1.8` aufgelistet.

Requests

Die Startzeile eines Requests besteht aus der RTSP-Methode, die auf den Server, auf die Präsentation oder auf einen Strom angewendet werden soll, sowie aus der RTSP-URL

Startzeile	OPTIONS rtsp://194.6.181.114:554 RTSP/1.0
Kopfzeilen	CSeq: 1 User-Agent: RealMedia Player (HelixDNAClient)/10.0.0.0 (win32) ClientChallenge: e42c674f5d812ae1ad7993e4508bea44 ClientID: WinNT_5.1_10.0.0.0_play32_RN01_EN_UNK CompanyID: AvbfpsajBKru5L+ipnoyPQ== GUID: 00000000-0000-0000-0000-000000000000 PlayerStarttime: [03/11/2004:19:26:32 01:00] Pragma: initiate-session RegionData: 0

Abbildung 1.6: RTSP-Request des Clients

Startzeile	RTSP/1.0 200 OK
Kopfzeilen	Date: Wed, 03 Nov 2004 18:26:30 GMT Session: 29809-1 Server: Helix Server Version 9.0.3.916 (sunos-5.8-sparc-server) (RealServer compatible) Public: OPTIONS, DESCRIBE, ANNOUNCE, PLAY, SETUP, GET_PARAMETER, SET_PARAMETER, TEARDOWN RealChallenge1: 51e1300382eb2cdcd96d73f99a53acf9 StatsMask: 3

Abbildung 1.7: Response des Servers

und der Protokollversion. Die einzelnen Werte sind dabei jeweils durch ein Leerzeichen getrennt. Der Request `DESCRIBE rtsp://media.server.com/foo/example RTSP/1.0` enthält beispielsweise die Anfrage an den Server, die Methode DESCRIBE auf die Präsentation /foo/example zu anzuwenden.

Wird anstelle einer URL der Stern * angegeben, so bezieht sich die Methode auf den Server selbst und nicht auf eine Präsentation oder einen Strom. Ein Bespiel für die Anwendung dieser Syntax ist die Methode OPTIONS: `OPTIONS * RTPS/1.0`

Um aufeinander folgende Anfragen und Antworten einander zuordnen zu können, enthält jede Anfrage eine CSeq-Kopfzeile, in der eine Sequenznummer enthalten ist. Die Antwort auf eine Anfrage enthält ebenfalls eine CSeq-Kopfzeile mit der identischen Sequenznummer.

Responses

Die Antwort auf eine RTSP-Anfrage besteht in der ersten Zeile aus der Protokollversion (RTSP/1.0) und einem Statuscode gefolgt von einer kurzen Beschreibung des Codes.

OPTIONS	Liefert eine Liste der auf dem Server verfügbaren Methoden.
DESCRIBE	Fordert die Beschreibung einer Präsentation oder eines Medienobjektes (identifiziert durch die URL) vom Server an.
ANNOUNCE	Aktualisiert die Beschreibung einer Präsentation oder eines Medienobjektes.
SETUP	Der Client liefert dem Server eine Liste mit vom Client unterstützten Transportparametern wie Transportprotokoll, Bandbreite etc. Der Server wählt daraus die für ihn optimale Variante aus und bestätigt sie dem Client.
PLAY	Fordert den Server auf, die Übertragung des Medienstromes mit den in SETUP definierten Transportparametern zu beginnen. Ein Client darf erst eine PLAY-Anfrage senden, wenn er eine erfolgreiche Bestätigung der SETUP-Anfrage vom Server erhalten hat.
RECORD	Startet die Aufnahme auf dem Server
PAUSE	Unterbricht die Präsentation oder einen einzelnen Strom vorübergehend (z.B. Audiostrom: Stummschalten).
REDIRECT	Der Server informiert den Client, dass er sich mit einem anderen Server verbinden muss.
GET_PARAMETER	Liefert einen Parameter, der einer Präsentation oder einem Strom zugeordnet ist.
SET_PARAMETER	Setzt den Wert eines Parameters einer Präsentation oder eines Stromes auf einen angegebenen Wert.
TEARDOWN	Beendet die Datenübertragung des Stromes und gibt die dafür allozierten Ressourcen frei. Wird als Parameter eine Präsentation angegeben, so wird die Sitzungs-Identifizierung dieser Sitzung ungültig.

Abbildung 1.8: Von RTSP unterstützte Methoden

Die weiteren Zeilen sind der Nachrichtenkörper, welcher beliebigen Text enthalten kann. Die Statuscodes sind dreistellige Zahlen, wobei die vorderste Ziffer die Klasse des Codes definiert:

- 1xx Information
- 2xx Erfolg
- 3xx Umleitung
- 4xx Client Fehler
- 5xx Server Fehler

Parameter

- URL Eine RTSP-URL kann auf eine Präsentation oder auf einzelne Medienströme verweisen. Dabei wird folgende Form verwendet:
  rtsp://media.server.com/foo/video
  Dieses Beispiel verweist auf den Strom video innerhalb der Präsentation foo, welche auf dem Server media.server.com abgelegt ist. Als Protokollschema wird rtsp:// für eine Kontrollverbindung über ein zuverlässiges Protokoll (im Internet TCP)

und rtspu:// für Kontrollverbindungen über ein unzuverlässiges Protokoll (UDP) verwendet. Falls kein Port explizit angegeben wird, so wird der Standardport 554 [18] verwendet.

- Identifiers Zur eindeutigen Identifizierung von Konferenzen und Sessions verwendet RTSP zwei Identifiers. Die Konferenz-ID, welche global eindeutig sein muss, sowie die Session-ID, welche zwischen Client und Server eindeutig sein muss. Beide bestehen aus alphanumerischen Werten und werten nach dem URI-Standard [40] codiert.

- Zeitstempel Die Position innerhalb der Präsentationen und Medienströmen wird mit Zeitstempeln angegeben, wobei drei Arten möglich sind: Der SMPTE Relative Timestamp, welcher die Zeit relativ zum Anfang des Clips angibt, die Normal Play Time, welche die Position relativ zum Beginn der Präsentation als Dezimalzahl in Sekunden angibt sowie ein Zeitstempel im ISO 8601 Format [19] zur Angabe der absoluten Zeit.

Implementierungen

Für RTSP existieren verschiedene Implementierungen, wobei meist eine Streamingserverplattform zusammen mit einem entsprechenden Player auf der Clientseite angeboten werden, welche so aufeinander abgestimmt sind, dass optimierte herstellerspezifische Formate, beispielsweise für DRM, verwendet werden können.

Die folgende Aufzählung gibt einen Überblick der drei momentan weit verbreiteten kommerziellen Produkte sowie von Open Source Software.

Server Für RTSP existieren verschiedene Implementierungen, wobei meist eine Streamingserverplattform zusammen mit einem entsprechenden Player auf der Clientseite angeboten werden, welche so aufeinander abgestimmt sind, dass optimierte herstellerspezifische Formate, beispielsweise für DRM, verwendet werden können. Die folgende Aufzählung gibt einen Überblick der drei momentan weit verbreiteten kommerziellen Produkte sowie von Open Source Software.

- Real Networks entwickelte die im April 1998 freigegebene Streaming Media Platform RealSystem G2 [27]. Dies war eines der ersten Medienserver-Systeme, in dem RTSP implementiert wurde. Im Januar 2003 wurde von RealNetworks der Quellcode von Helix DNA Server unter die Open Source Lizenz gestellt. Helix [10] ist eine Serverplattform, die das Erstellen und Übermitteln von Multimediainhalten in Echtzeit ermöglicht

- Apple Computer entwickelte den Quicktime Streaming Server, der im April 1999 veröffentlicht wurde. Mittlerweile bietet Apple den Darwin Open Source Streaming Server [4] an, der die Protokolle RTSP und RTP verwendet, um Quicktime und AVI-Videoströme zu übertragen.

- Ein weiteres Open Source Projekt ist die C++-Bibliothek Live.com Streaming Media [23]. Sie unterstützt ebenfalls RTSP und RTP/RTCP und ermöglicht ausserdem das Senden, Empfangen und Verarbeiten von MPEG Videos.

- In der Microsoft Windows 2003 Serverplattform sind die Windows Media Services [44] enthalten, welche nebst meist proprietären Formaten auch RTSP unterstützen.

Clients Die drei grossen Hersteller von Streaming-Serverplattformen Apple, Microsoft und Real Networks bieten auch je einen RTSP-fähigen Medienplayer zum kostenlosen Download an, welcher für das vom Hersteller jeweils verwendete, meist zumindest teilweise proprietäre Streamingformat, sowie für den herstellerspezifischen Codec optimiert ist. Die hier erwähnten Player unterstützen die Wiedergabe von Audio- und Videostreams sowie zusätzlich die Wiedergabe von weiteren Codecs wie MP3-Audio und MPEG-Video. Ausserdem können sie auch als Plugin in Webbrowser wie Mozilla, Internet Explorer oder Opera eingebunden werden, so dass die Wiedergabe der multimedialen Inhalte direkt im Browser erfolgen kann. Nebst der eigentlichen Hauptfunktion, dem Wiedergeben von Audio- und Videostreams, bieten die erwähnten Player auch noch eine Reihe von Zusatzfunktionen wie das Brennen von Multimediadaten auf CD, die Möglichkeit zum kostenpflichtigen Download von Musikdateien oder das Suchen nach Webradiostationen. Die momentan am häufigsten eingesetzten Clients sind der Real Player von RealNetworks, der Windows Media Player von Microsoft und Apples Quicktime Player.

Neben Client-Anwendungen für Endbenutzer steht den Entwicklern von Streaming-Anwendern das Java Media Framework zur Entwicklung eigener Lösungen zur Verfügung: Das von von Sun Microsystems, Silicon Graphics, Intel und IBM entwickelte Java Media Framework [20] ermöglicht die Entwicklung von RTSP-Clients in Java.

1.3 Kommerzialisierung von Real-Time Data Transfer im Internet

1.3.1 Entwicklung des e-Commerce (Internet Economics)

"Now I assign the Internet the highest level of importance" - dieser Satz aus dem internen Memo von Bill Gates aus dem Jahre 1995 zeigt die Bedeutung dieser digitalen Infrastruktur, welche in zunehmendem Masse als Marktplatz genutzt wurde. Schnell erkannte man auch, dass dieser virtuelle Marktplatz dem bekannten Prinzip der Ökonomie von Angebot und Nachfrage unterliegt (vgl. Zerdick et al. 2001 [45]). "Der grundlegende Unterschied besteht darin, dass ein Teil der oder die gesamte Transaktion mit Hilfe von entsprechenden Informations- und Kommunikationssystemen elektronisch abgebildet und abgewickelt wird" [45].

Die fortschreitende Entwicklung der Informations- und Kommunikationstechnologien vor allem in Bezug auf das Internet bis hin zur Unterstützung und teilweisen Automatisierung

aller Ebenen des kommerziellen Handels ermöglichen die Realisierung neuer Marktpotentiale, deren verschiedene Formen unter dem Schlagwort 'e-Commerce' zusammengefasst werden. Beim e-Commerce kommt die ganze technologische Bandbreite des Internets zum tragen und ermöglicht, vollständige Geschäftsabläufe, von der Information (Awareness) über die Bestellung (Transaction) und die Logistik bis hin zur Kundenbetreuung (Customer Relationschip Managment) elektronisch abzuwickeln. E-Commerce befähigt mit der Eröffnung einer weltweiten Filiale (Internet) zur internationalen Markterschliessung zu relativ geringen Kosten. Das weitreichende Nutzenpotential der digitalen Geschäftsabwicklung im Internet in Bezug auf Absatzpotential, Kundenorientierung, Zeit- und Kosteneinsparungen, Positionierung im Wettbewerb sowie der Realisierung neuer Erlösmodelle wurde schnell erkannt und ist die treibende Kraft der Technik (vgl. Zerdick et al. 2001 [45]).

1.3.2 Prämissen und Chancen

Eine der Grundlagen für Streaming-Media ist sicherlich, das Umfeld in dem es sich befindet. Bereits bestehend und am Markt durchgesetzt ist der allgemeine Rundfunk von Radio und Fernsehen (Broadcast), sowohl von öffentlich-rechtlichen wie auch von privaten Stationen respektive Unternehmen. Ein weiterer Bereich lässt sich sicherlich im Verkauf und Verleih von Medien-Datenträgern wie DVD oder Musik-CD ansiedeln, der wiederum im Gegensatz zur neuen technischen Entwicklung bereits im Markt integriert ist und sich demnach in einer Konkurrenzbeziehung zu den Streaming-Media Angeboten befindet.

Eine weitere Voraussetzung von Streaming-Media ist die technische Grundlage. Die Entwicklung von Streaming-Media basiert auf der Technik, worauf wir später unsere Argumentation aufbauen. In diesem Zusammenhang spricht man auch von einem Technikdeterminismus (Hennen 1992 [12]).

In den Anfängen war es undenkbar Sprache, Musik oder gar Videos über das Internet zu übertragen. Mit der Entwicklung von Komprimieralgorithmen für Audio- und Videodaten, vor allem dem MPEG-Verfahren, war es erstmals möglich, diese ungeheuren Datenmengen auf eine Internet-taugliche Größe zu bringen, welche für den Endbenutzer dementsprechend einfach und qualitativ hochstehend darzustellen war. Der gemeinsame Einsatz von Streaming Protokollen und Audio- bzw. Videokompression führte schließlich zum Audio- und Videostreaming und es ergaben sich dadurch ungeahnte neue Möglichkeiten. Es handelt sich hierbei um einen Übergang zur digitalen, nicht physischen Produktion und Distribution von Informationsinhalten. "Während traditionell Information eindeutig und fest an bestimmte Medien gekoppelt war, lassen sich nunmehr ein und dieselben Inhalte über unterschiedliche Medienkonfigurationen verbreiten. Diese zunehmende Trennung von Medium und Information wird im Weiteren als Desintegration bezeichnet".
(Zerdick et al. 2004 [46]) Wichtigster Bestandteil der Desintegration des Mediums ist, dass die Koppelung der Information an Medien sich zum Endkunden hin verschiebt und den Übergang vom traditionellen 'Marketplace' in den 'Marketspace' veranlasst. In der Konsequenz kann die Wertschöpfungskette immer weniger physisch werden, und erhebliche Vorteile für Customer und Business mit sich bringen. Den Anbietern von Content wie TV, Radio und Printmedien ermöglicht es eine weitere Markterschliessung, neue Geschäfts-

und Ablösemodelle sowie die Vereinfachung der Prozesse, da es am Beispiel einer Online Zeitung möglich ist, die gesamte Wertschöpfungskette digital und nicht mehr physisch zu verarbeiten.

Für die Endkunden gibt es mehrere Arten an Information zu gelangen und diese entsprechend zu nutzen. Es ist dem Endkunden überlassen ob und in welcher Form er die übertragenen Inhalte erstellt und nutzt. Die Verlagerung der Wertschöpfungsprozesse in die immaterielle Ebene der Information ist kein Spezifikuum der Medienindustrie und geht zurück auf die Erkenntnisse der Wissenschaftler Rayport und Sviokla für alle Branchen aus dem Jahre 1994. Nichtsdestotrotz vollzieht sich dieser Wandel in der Medienbranche umso radikaler, da sich zum einen Information leicht digitalisieren lässt und zum anderen der Kundennutzen letztlich durch die digital transportierten Inhalte entsteht und nicht durch das physische Medium (vgl. Zerdick et al. 2004 [46]).

1.3.3 Streaming-Media als Innovation

Gründe für ein Unternehmen sowie für einen Endnutzer, sich mit den Möglichkeiten und Angeboten im Streaming-Media Bereich zu konfrontieren und auseinander zu setzen, gibt es viele, jedoch fallen diese teilweise auch negativ ins Gewicht. Um Aussagen über die kommerzielle Potenz von neuen Angeboten wie Streaming-Media machen zu können, ist es notwendig, das angebotene Medium in all seine Facetten zu zerlegen. Im Weiteren werden wir die aktuellen Angebote im deutschsprachigen Raum analysieren, und versuchen Anhand deren Diffusion und Nutzung die wesentlichen Merkmale sowie die Relevanz von Streaming-Media und somit die wirtschaftliche Kraft festzuhalten.

Streaming-Media ist mitunter ein technischer Wandel in anbetracht der Herstellung (Content Provider, Production, Format), Verbreitung (Awareness, Distribution, Transport, Logistic) und Nutzung (Customazation, Customer Relationschip Management) des Mediums (Video und Audio).

Ein technischer Wandel lässt sich nach Kiefer in einem Dreischritt umschreiben. Invention, Innovation und Diffusion. Invention kann definiert werden als „grundlegende Entdeckung “eines zuvor nicht wahrgenommenen technischen und ökonomischen Zusammenhangs oder einer entsprechenden Handlungsmöglichkeit (Kiefer 2003 [21]). Innovation dagegen sind Neuerungen, die sich am Markt durchsetzten (vgl. Hotz-Hart/Reuter/Vock 2001 [13]), wobei Letzteres für Streaming-Media noch unerforscht ist. Unsere Beobachtungen und Analyse der aktuellsten Angebote von Streaming-Media im deutschsprachigen Raum, welche unter 1.3.6 aufgezeigt sind, lassen vermuten, dass diese neue Art von Angebot sich noch in den Kinderschuhen befindet.

Arten von Innovationen

Innovationen können auf mehreren Ebenen unterschieden werden. Die erste betrifft die Art der Innovation. Zielt die Innovation auf ein Produkt ab, spricht man von Produktinnovationen. Produktinnovationen beschreiben sowohl neue Produkte wie auch Innovationen

in der Herstellung bereits bestehender Produkte. Weitere Arten von Innovationen sind Organisationsinnovationen. Sie bezeichnen die Veränderungen in der Koordination des arbeitsteiligen Leistungserstellungsprozesses, bei den unternehmerischen Standorten und im Management (vgl. Hotz-Hart/Reuter/Vock 2001 [13]).

Eine zweite Unterscheidungsebene kann gefunden werden, wenn man Innovationen betreffend ihrer volkswirtschaftlichen und gesellschaftlichen Auswirkungen betrachtet. Radikale oder Basisinnovationen sind tiefgreifende Innovationen, die Innovationsschübe auslösen und zu erheblichen Strukturveränderungen führen. Inkrementelle Innovationen bezeichnen dagegen Innovationen, welche nur eine begrenzte innovationsökonomische Wirkung haben (vgl. Hotz-Hart/Reuter/Vock 2001 [13], Kiefer 2003 [21]).

Weiter können disruptive von erhaltenden (sustaining) Innovationen unterschieden werden. Disruptive Innovationen bezeichnen neue Produkte, die in einen bereits bestehenden Markt eingeführt werden. Typisch für diese Innovationen sind schlechtere Produktqualität, der tiefere Preis und auch die Vereinfachung des Produktes. Mit der Weiterentwicklung dieser Produkte übertreffen diese Produkte die bewährten mit der Zeit qualitativ und drängen diese aus dem Markt. Erhaltende Innovationen dagegen sind Produkte, die keine wesentlichen Änderungen beinhalten, aber eine bessere Performance erzielen oder zu einem günstigeren Preis angeboten werden (vgl. Christensen 1997 [3]).

Gemäss diesen Definitionen kann Streaming Media klar zu den disruptiven Innovationen gezählt werden. Der bestehende TV Markt über Kabel und Satellit ist im deutschsprachigem Raum weit verbreitet. Streaming Media verfügt nach wie vor über eine schlechtere Produktqualität, ist vielerorts billig bis gratis verfügbar und lässt sich als generelle Vereinfachung des TV's einstufen.

Gesellschaftliche und Wirtschaftliche Bedeutung von Innovationen am Beispiel von Streaming Media

In der Art und Weise, wie Technik und insbesondere Medien gesellschaftlichen Wandel determinieren, unterscheiden sich die verschiedenen Ansätze stark. Allen ist allerdings gemeinsam, dass nicht der Inhalt der Medien, sondern deren technische Form und die damit verbundenen Sachzwänge entscheidend sind. Einige Autoren versuchen, anhand von historischen Veränderungen der Medien epochale Wandel der menschlichen Geschichte zu begründen (vgl. Innis 2003 [17]). Andere erklären handlungstheoretisch, dass durch die zunehmende und autonome Technisierung sich zweckrationales und der Technik angepasstes Handeln zum dominanten gesellschaftlichen Orientierungstypus entwickelt (vgl. Habermas 1981 [7]). Weitere Ansätze betrachten technische Medien als Organe zur Wahrnehmung der Umwelt und als Instrument der Wirklichkeitserzeugung. Durch den technischen Wandel wird so die Vermittlung von Realität wesentlich geprägt (vgl. Hartmann 2003 [9]). Technikkritische Vertreter gehen von einer sehr starken Determinierung der Gesellschaft durch Technik aus, so dass sie im technischen Wandel die absolute Steuerung der Gesellschaft sehen. Demokratische Gesellschaften würden deshalb durch die zunehmende Macht der Technik zum "Technopol" hin tendieren (vgl. Postman 1992 [26]).

Folgendes Zitat verdeutlicht aber im Sinne einer Kritik zur technikdeterminierenden Ansicht, dass man ebenso von einer sozialdeterminierenden Perspektive ausgehen kann: *"Die*

Auffassung der Technik als soziokulturell unabhängige Variable, als zweckrationales Instrument der Kontrolle der Umwelt, trifft sicherlich einen wichtigen Aspekt von Technik. Technische Artefakte sind immer Instrumente der Kontrolle der Umwelt, der Entlastung und Erweiterung von Handlungsmöglichkeiten. Sie sind dies aber nicht unabhängig vom Wissen gesellschaftlicher Akteure: Sie sind Produkte menschlichen Handelns und sind als Instrumente wieder in soziale Handlungskontexte eingebettet“ (Hennen 1992 [12]).

Somit kann man je nach Definition zur einen oder der anderen Ansicht tendieren. Wir positionieren uns in dieser Arbeit in der Mitte, betrachten wir doch auf Grund der gesellschaftlichen Veränderung durch die Technik und das soziale Handeln, dass beide Ansätze ihre Richtigkeit besitzen. Des Weiteren lebt die wirtschaftliche Potenz ebenfalls von diesem Zusammenspiel der beiden Ansätze, da die Technik-Industrie die Endgeräte so einfach und nützlich machen muss wie möglich, und doch das Gerät im Endeffekt durch das Handeln der Endnutzer bestimmt wird, sowie versteckte Verbesserungsbotschaften in Form von Akzeptanz, Relevanz und Feedback das Gerät weiterentwickeln können.

Mit einer radikalen Basisinnovation wie die Entwicklung der RTP und RTSP Protokolle und schnellen Netzwerken werden neue technische Paradigmen geschaffen, innerhalb derer sich dann eine Vielzahl von inkrementellen Innovationen bzw. Innovationscluster entwickeln. Neue radikale Innovationen (neuer technischer Standard) können dann wieder zu neuen technischen Paradigmen führen (vgl. Dosi 2000 [6]).

Innovationen sind für jede Marktwirtschaft zentral. Jede ökonomische Entwicklung baut auf dem Prozess der schöpferischen bzw. kreativen Zerstörung auf. Durch die Zerstörung von alten Strukturen werden Produktionsfaktoren immer wieder neu geordnet, somit ist sie notwendig, damit Neuordnung stattfinden kann. Dieser Prozess der schöpferischen Zerstörung kann als industrielle Mutation bezeichnet werden. Auslöser der schöpferischen Zerstörung sind Innovationen.

Dem gegenüber findet man in der Literatur Wolfgang Riepl's Unverdrängbarkeitsgesetz wonach “Medienangebote nie gänzlich und dauerhaft verdrängt und ausser Kraft gesetzt werden [...], sondern sich neben diesen erhalten, nur dass sie genötigt werden, andere Aufgaben und Verwertungsgebiete aufzusuchen“ (Riepl 1912, zit. nach Hagen 2002 [8]).

Auch für die Gesellschaft haben Innovationen eine hohe Bedeutung. Auf der einen Seite erleichtern sie dem Menschen das Leben, auf der anderen Seite ist jeder Einzelne darauf angewiesen, sich den neuen Gegebenheiten anzupassen um nicht gesellschaftlich isoliert zu werden (vgl. Hotz-Hart/Reuter/Vock 2001 [13]). Dies wiederum birgt Gefahren in der Gesellschaft, welche in der Wissenschaft unter dem Schlagwort 'Digital Divide' heftigst diskutiert werden. In den vergangenen Jahren fand in der Öffentlichkeit eine umfangreiche Auseinandersetzung statt, die vermehrt auch im gesellschaftlichen Kontext an Relevanz gewinnt. Es geht dabei um eine zentrale Ressource in der Informationsgesellschaft, nämlich dem Zugang zu Information und deren Nutzung. Neue Medien wie Internet spielen hierbei eine wichtige Rolle, zumal sie den Zugang als auch die Nutzung von Informationen erheblich vereinfachen können. Der immer kurzlebigere Innovationszyklus und die damit verbundenen Anschaffungskosten für den Zugang zu neuen Medien bringen es mit sich, dass gewisse Teile der Bevölkerung nicht oder nur eingeschränkt mit dieser Entwicklung Schritt halten können und wollen. Nicht Jedermann ist gewillt und hat die nötigen finanziellen Mittel mit dieser Dynamik stetig einherzugehen. Somit lässt sich erahnen, dass die

Diffusion des Internets sozial sehr heterogen sein kann. Stark überspitzt formuliert könnte man sagen; die Gesellschaft spaltet sich vor dem Hintergrund der Zugangskluft in eine technologieaffine Informationselite und in eine technologieabstinente Informationsmasse. Dabei ist die Perspektive auf der ersten Ebene (First Level Digital Divide) ausschliesslich auf Zugang/Nicht-Zugang gerichtet und blendet weitere Unterschiede vollends aus, während auf der zweiten Ebene (Second Level Digital Divide) nur die unterschiedliche Nutzung innerhalb der digitalen Medien (Internet) betrachtet wird (vgl. Marr 2003 [24]).

Kommunikationstechnologische Bedeutung von Streaming Media

Der Kommunikationssektor ist stark von neuen Protokollen als Innovationen wie RTP und RTSP geprägt. Medien sind technische Interaktionssysteme, die ohne Technologie gar nicht denkbar wären. Die Kommunikation von Inhalt und Information ist eine Form technisch übermittelter Kommunikation von 0 und 1, die Raum und Zeit überwindet (vgl. Seeger 1999 [33]). Die Entwicklung von Innovationen geht dabei fast nicht von dem Kommunikationssektor aus: typisch für diesen Sektor ist, dass Innovationen von anderen Wirtschaftssektoren übernommen werden und allenfalls betriebsspezifische Adaptionen vorgenommen werden (vgl. Kiefer 2003 [21]). Die Informatik-Produkte (Protokolle) wurden demnach von der Kommunikationsbranche (Rundfunkbetreiber) übernommen und für die Entwicklung eigener Geschäftsmodelle angepasst und eingeführt. Die Übertragungs-Protokolle (Kommunikation) werden mit Kommunikationsgütern (Information) in Verbindung gebracht und daraus ergeben sich neue Absatzmodelle und Möglichkeiten von zwei Gütern, die in einem kommerziellen Zusammenhang gebracht wurden. Anwendungsbeispiele sind weiter in 1.3.7 aufgeführt Die technische Geschichte der Medien lässt sich ökonomisch auch als eine Geschichte der Auflösung des Uno-actu-Prinzips begreifen. Das Uno-acto-Prinzip bedeutet, dass Produktion und Konsum örtlich und zeitlich zusammenfallen. Die Auflösung dieses Prinzips bedeutet, dass der Rezipient Medien zu jedem Zeitpunkt und an jedem Ort konsumieren kann (vgl. Kiefer 2003 [21]). Genau dieser strategische Gedanke ist für die Kommerzialisierung von Real-Time Data Transfer von grosser Bedeutung, ermöglicht er doch die zeitliche und örtliche unabhängige Rezeption von Informationsinhalten.

Substitution vs. Komplementarität

Unter Substitution versteht man die ökonomisch, zeitlich oder funktional motivierte Ersetzung der Nutzung eines Medienangebotes durch die Nutzung eines Anderen. Sind die Gratifikationen zweier Angebote nahezu deckungsgleich, so entscheidet sich der Rezipient für dasjenige, welches mit dem geringsten Kosten- und Zeitaufwand verbunden ist (vgl. Trepte/Baumann/Borges 2000 [38]).

Komplementarität hingegen liegt dann vor, wenn mindestens zwei Medienangebote ergänzend genutzt werden, weil die Nutzung des einen Angebotes die des anderen voraussetzt oder bedingt.

1.3.4 Diffusion und Relevanz

Um einen weiteren theoretischen Aspekt für die Diskussion von 1.4 als Grundlage zu haben, möchten wir die Erkenntnisse der Diffusionsforschung aufgreifen. Eine dem Diskurs 'kommunikationstechnischer Wandel' zu Grunde liegende Theorie ist diejenige der Diffusionsforschung. Diese befasst sich hauptsächlich mit der Schaffung und Verbreitung grundlegender technologischer Innovationen, wobei meistens von der Synthese des Wissenschaftlers Everett M. Rogers ausgegangen wird. Rogers beschreibt die am Verbreitungsprozess ausschlaggebenden Faktoren (knowledge / persuasion / decision / implemetation / confirmation) für die Diffusion einer Innovation (vgl. Rogers 1995 [31]).

"There is nothing more difficult to plan, more doubtful of success nor more dangerous to manage than the creation of new order of things [...]" Rogers 1995 [31]. Nichtsdestotrotz konnte Rogers aufzeigen, dass die Diffusion von Innovationen, die Durchsetzung neuer Technologien, immer einem typischen Muster folgt. Die Akzeptanz bzw. die Verweigerung einer Innovation kann, ähnlich wie bei McLuhan, zu einer Umgestaltung in der Gesellschaft führen (vgl. Rogers 1995 [31]). Dieses 'typische Muster' wird in Form einer S-Kurve wiedergegeben (vgl. `Abbildung 1.9`).

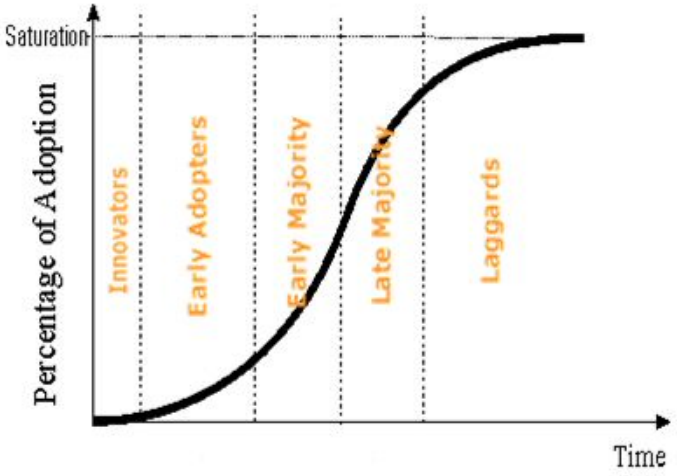

Abbildung 1.9: Saturationskurve

Dabei variiert je nach Diffusion die Kurve in der Länge der Zeit die die Innovation benötigt, um zum Saturationspunkt zu gelangen. Die Kurve fällt demnach steiler oder flacher aus. Innerhalb dieses Prozesses kann man verschiedene Arten von Gruppen ausmachen, die sich voneinander durch den Zeitpunkt der Adaption einer Innovation unterscheiden. Diese Gruppen werden in anderen Studien oft mit demographischen Merkmalen umschrieben, worauf wir hier in unserer Arbeit jedoch verzichten, da es den inhaltlichen wie auch zeitlichen Rahmen sprengen würde. Wichtig fest zu halten ist, dass der Adaptionsprozess von der Kenntnisnahme (knowledge), über die Überzeugung (persuasion), der Entscheidung (decision), bis hin zur eigentlichen Implementierung (implementation) und der Bestätigung einem bestimmten Muster folgt, welches jedoch auch durch den Markt oder der Industrie verändert werden kann (Rogers 1995 [31]).

1.3.5 Technology push vs. Market / Demand pull

Gemäss dem Wissenschaftler Dosi kann nun diese Kurve massiv durch Markt oder Industrie beeinträchtigt werden. Wie in 1.10 zu sehen ist, wird teils durch die Nachfrage des Marktes die Kurve zu einem gewissen Zeitpunkt steiler. Dies ist vor allem der Fall, wenn sich die neue Innovation als sehr gut erwiesen hat und teils bereits als das Standardgerät gilt. Andere Arten von Market-Pull, können sich auch auf Grund der Topologie der Nutzung der Innovation herauskristallisieren. So sind die bekannten Tauschbörsen auf dem Internet gerade so gut wie die Anzahl ihrer Nutzer. Dosi spricht in diesem Zusammenhang von einem Netzeffekt. Je mehr Nutzer der Technologie es hat umso qualitativ und quantitativ hoch stehender wird die Anwendung (vgl. Zerdick et al. 2001 [45]).

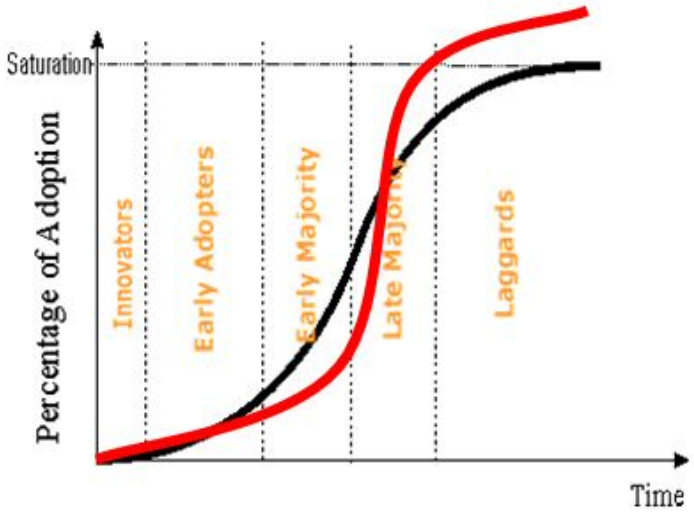

Abbildung 1.10: Interaktive Diffusionskurve

Auf der anderen Seite werden oft auch Produkte durch die Industrie gefördert, wenn man beim Saturationspunkt einer Innovation angekommen ist. Dies sind vor allem Produkte für die es im eigentlichen Sinne gar keinen Markt geben dürfte. So wurden Beispielsweise Wide-Screen Fernseher angeboten, bevor irgendwelche dementsprechend codierte Filme auf dem Markt waren. Diese Effekte kann man auch als so genannte Lock-In Effekte beschreiben, da solche technische Neuerungen jeweils nur mit ebenfalls neuen spezifischen Peripherie-Geräten funktionieren oder ihre neue Funktion nur mit Hilfe von Drittgeräten ersichtlich werden (vgl. Dosi 1982 [5]).

1.3.6 Aktueller Streaming-Media Markt

Bevor wir nun uns auf den bestehenden Markt stürzen, möchten wir noch aktuelle Internetnutzungsdaten der Schweiz aufzeigen, um den Rahmen der Relevanz aufzuzeigen, in dem wir uns mit der Streaming-Media befinden und konfrontiert sind.

Gemäss der Fernmeldestatistik des Bundesamt für Kommunikation im Jahre 2003 [2] haben 2.7 Mio der Schweizer Bevölkerung einen privaten Internetanschluss. Lediglich 27%

dieser Internetnutzer verfügen jedoch über einen für Streaming-Media Angebote erheblich wichtigen Breitband-Anschluss, was insgesamt 10% der Gesamtbevölkerung ausmacht.

Laut der MA-Comis der AG für Werbemedienforschung 2004 [43] gibt es in der Schweiz insgesamt 3.8 Mio regelmässige Internetnutzer, sprich fast jede zweite Person benutzt das Internet in regelmässigen Abständen an öffentlichen und privaten Anschlüssen. Eine eigene Fortschreibung und Untersuchung dieser Daten aus demselben Jahr, ergab, dass über 80% der Internetnutzer das Senden und Empfangen von E-Mails benutzen. Rund 65% benutzen Suchmaschinen und lediglich jede 5 Person (20%) benutzt Angebote im Streaming-Media Bereich (vgl. WEMF [43], eigene Fortschreibung). Diese Zahl ist umso imposanter wenn man bedenkt, dass die Kommerzialisierung von Streaming-Media noch ziemlich in den Kinderschuhen steckt. Das `nachfolgende Abbildung 1.11` bietet einen Überblick im Streaming-Media Bereich und zeigt auf, wie unterschiedlich die Protokolle RTP/RTSP genutzt werden können:

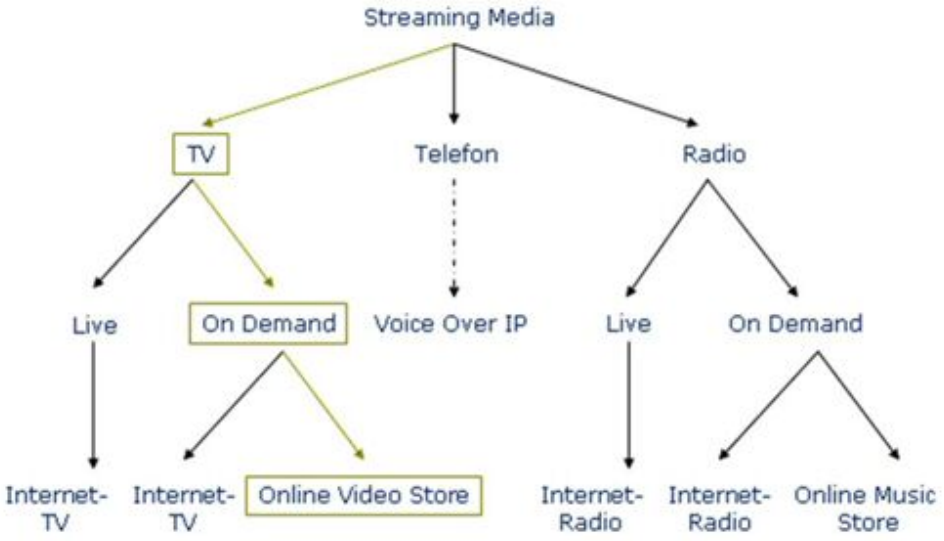

Abbildung 1.11: Streaming-Media Angebote

Nebst der Unterscheidung von TV, Radio und Telefon, werden die Angebote auf ein weiteres wesentliches Merkmal überprüft. Wir ziehen die Unterscheidung zwischen Live und Video On Demand (VOD) in Betracht, eröffnet sie doch eine weitere Ebene der Markterschliessung und Marktpositionierung. Welches Format sich als Substitutions- oder Komplementärfunktion definieren lässt, werden wir in der Diskussion wieder aufgreifen (vgl. `Kap 1.4`).

1.3.7 Anwendungsbeispiele

Der Bereich von Streaming-Media ist enorm gross, praktisch so viele kleine, und regionale TV-Stationen stellen ihr volles Programm Live auf das Internet und machen es einem weltweiten Publikum zugänglich. Bei praktisch allen Versuchen sich in den Live Stream

der TV Stationen einzuschalten, wurden jedoch technische und qualitative Mängel bemerkbar. Vielerorts war die Qualität der Wiedergabe so schlecht, dass man weder Bild noch Ton ausmachen konnte. Die meisten öffentlich-rechtlichen wie auch grössere private Fernsehsender stellen nur eine begrenzte Auswahl ihrer Sendungen auf das Netz und diese verstehen sich als Video On Demand-Anbieter. Eine Ausnahme macht hier der deutsche Fernsehsender ZDF [55] (öffentlich-rechtlich) und der Musiksender VIVA [53] (privat), welche einige Sendungen auch als Live-Stream zur Verfügung stellen (wwiTV [54]). Das Schweizer Fernsehen DRS präsentiert eine Auswahl ihrer Sendungen als Video on Demand und verweist auf ein enormes Archiv, welches beispielsweise bei der Sendung ins Jahr 2000 zurückgeht. Ein beachtliches Engagement, wenn man beachtet, dass diese Sendung fünfmal wöchentlich ausgestrahlt wird. Hierbei kann der User ebenfalls die Qualität bestimmen indem er zwischen 'Low' und 'High' auswählt. 'Low' wird mit einer Bitrate von 80 kbit/s übermittelt, während 'High' mit zwischen 318 kbit/s und 350 kbit/s übermittelt wird. Durch den RTSP Standard lassen sich die Sendungen beliebig vorwärts und rückwärts 'spulen' (vgl. SFDRS.ch [52]). Dieses enorme Sortiment an Video on Demand vom öffentlich rechtlichen TV- Sender versteht sich als eine Gratisdienstleistung. Dies ist allerdings nicht die Norm; Das österreichische Staatsfernsehen ORF [48] zum Beispiel bietet keinerlei Streaming Video an. Bei den Privatsendern findet man bei RTL [50] und Pro7 [49] ebenfalls keinerlei Streaming Video, lediglich der Internetauftritt von Sat1 [51] bietet von einigen seiner Sendungen einen Abspann (Trailer/Vorschau) an. Als Gegenbeispiel möchten wir an dieser Stelle noch eine weitere Art von Streaming-Angeboten aufzeigen. Die private Fernsehanstalt CNN [47] in den USA verspricht die Aufschaltung von bis zu 30 neuen Videos pro Tag und bietet viele Sendungen als Video on Demand an. Jedoch muss man dabei ein Abonnement gelöst haben. Der Zugang ist in den ersten 14 Tagen gratis und wird danach mit EUR 17.50/Monat verrechnet. Dieses B2C Modell wird bei weiteren Firmen und Organisationen angewendet. Der europäische Fussballverband UEFA hat auf seinem Internetauftritt ein Video on Demand Programm bei welchem der User wiederum eine monatliche Grundgebühr von EUR 4.95 begleichen muss um auf Highlights aller Spiele der Champions League sowie auf Interviews und Teilaufzeichnungen mit Original-Kommentar zugreifen zu können.

Viele Städte, Flughäfen und Tourismusdestinationen haben in Ihrem Internetauftritt eine Live-Web-Cam integriert um einem Besucher Einblicke und Wetter-Informationen zu liefern (vgl. EarthCam [11]). Ein weiteres Feld von Video on Demand ist das Streaming von Kino-Filmen im Gegensatz zum eigentlichen TV-Programm der jeweiligen Fernseh-Stationen. Eine Art Video-Online Store, welcher Kino-Filme zum Streaming anbietet. Die verschiedenen Möglichkeiten im Bereich von Online-Stores erscheinen uns als sehr vielfältig und lukrativ, wobei die technologische Herausforderung für ein akzeptables Aufwand/Ertrag Verhältnis am Grössten ist. Die Betreiber dieses Video on Demand Angebots sind mehrheitlich Internet Service Provider (ISP) und betreiben die Angebote ausschliesslich gegen Entgelt. In `Abbildung 1.12` ist eine Übersicht dieser Angebote zu sehen.

1.4 Diskussion - The Innovators Dilemma

Diesen Begriff prägte der Wissenschaftler Christensen und umschrieb damit die Sackgasse in der sich führende Unternehmen bei der Einführung von neuen technischen Innovationen

Angebot	Abo / Monat	zusätzlich / Spielfilm	Nutzungs-dauer	Betreiber	Zugang	Besonderes	Registrier-ung	Bitrate
stream-it.ch	Fr. 5.-*	Fr. 1.- bis 7.-	24h	Bluewin.ch	nur für Bluewin ADSL-Kunden	Versuchs-betrieb	ja	?
Arcor Video-on-Demand	-	€ 2.50 / 3.50	24h	Arcor.de	nur mit deutscher IP zugänglich	-	ja	500kbit/s bis 1Mbit/s
T-Online Vision	-	€1.90 bis 4.00	24h	T-Online.de	nur für T-Online Kunden	-	Nein (Auto-Erkennung)	544kbit/s
one4movie.de	€ 9.95	- (Flatrate)	Un-beschränkt (innerhalb Abodauer)	4friends GmbH	Un-beschränkt	Flatrate	ja	?

Abbildung 1.12: Aktuelle Angebote - Online Video Stores im deutschsprachigen Raum

befinden (vgl. Christensen 1997 [3]). Soll man auf die neue Technik setzen, oder bleibt man dem alt bewährten treu? Insbesondere schwierig ist dieses Dilemma bei disruptiven Innovationen, wie dies in unserem Beispiel Streaming-Media der Fall ist, zu lösen. Als Beispiel sei hier der Telekommunikations-Anbieter Swisscom erwähnt, welcher sich im Festnetzbereich als Marktführer etabliert hat. Soll dieser Konzern nun auf das Voice Over IP-Geschäft setzen, mit dem Wissen, dass er sein eigenes Kerngeschäft konkurrenziert, oder soll er dieses Geschäft mit noch nicht erwiesenem Zukunftspotential anderen Marktteilnehmern überlassen?

Auf Grund solcher Fragen versuchen wir Vor- und Nachteile im Geschäft mit digitalen Medieninhalten herauszusuchen. Um hierbei möglichst präzise Beispiele geben zu können, werden wir lediglich im Bereich von Video-Online Stores (VOD), gemäss unserer Übersicht (1.12) argumentieren.

Wie wir bereits in der Einführung zum E-Commerce (vgl. 1.3.1) gesehen haben, liegt der grösste Vorteil der Digitalisierung bei der Überwindung von Zeit und Raum. Jeder kann wann und wo immer er möchte einen Spielfilm 'streamen'. Diese Entkoppelung der Inhalte vom Medium bringt erhebliche Vorteile in der Art und Weise wie und wo die Inhalte konsumiert werden können. Mit der enormen Datenmenge an Angeboten die den Internetnutzern zur Verfügung steht, steigt der Informations- und Unterhaltungswert einzelner Inhalte. Diese können nämlich wiederholt (On-Demand) gesehen werden und man hat theoretisch die Möglichkeit Hintergrundinformationen, Versprecher von Moderatoren usw., zu sehen, welche man vielleicht verpasst hat. Der grösste Nachteil von Streaming-

Media am Beispiel von Video On Demand ist sicherlich die benötigte Infrastruktur und die damit noch immer gekoppelte Kompetenz der Endbenutzer. Es gibt noch kein im Markt erprobtes Endgerät, welches an den bereits in den meisten Haushalten vorhanden Fernseher angeschlossen werden kann/wird und herkömmliches Fernsehen mit Video On Demand ergänzt. Des Weiteren scheint unserer Meinung nach, die mangelnde Qualität der Online-Angebote der massgebendste Punkt zu sein, warum die Leute es noch weiter vorziehen, in die lokale Videothek zu gehen um sich für relativ wenig Geld einen Film für 24 Stunden physisch auszuleihen. Die bei den Online-Angeboten möglichen Bit-Raten aus der Übersicht (vgl. `1.12`) liegen noch eine 10er Potenz unter denen des DVD Standards. Klar muss die Qualität nicht auf einem DVD Standard basieren, da es vielleicht in Zukunft extrem billig sein wird, Filme zu 'streamen', doch hier sind wir an einem kritischen Punkt in der Internet Ökonomie: Das Internet wird als inferiores Gut angesehen (vgl. Zerdick et al. 2004 [46]) wo niemand bereit ist zu zahlen. Wenn man zu zahlen bereit wäre, erwartet man dementsprechend eine hohe Qualität der erworbenen Güter und Inhalte, wovon wir im Streaming-Media Bereich noch weit entfernt sind. Die Angebote aus der Übersicht in Tabelle 2 waren zum Zeitpunkt der Erarbeitung dieser Seminararbeit die einzigen Angebote im deutschsprachigen Raum, und dabei handelte es sich teils noch um Testphasen mit beschränkten Teilnehmerzahlen (bluewin.ch). Unserer Ansicht nach steckt die gesamte Verarbeitung dieses Prozesses noch deutlich in den Kinderschuhen. Einer anderen Auffassung sind hierbei die Verfasser einer empirischen Arbeit der EMC Corporation über Real-Video: "Typical RealVideos have high quality, achieving an average frame rate of 10 frames per second and very smooth playout, but very few videos achieve full-motion frame rates. Overall video performance is most influenced by the bandwidth of the end-user connection to the Internet, but high-bandwidth Internet connections are pushing the video performance bottleneck closer to the server" (Wang/Zuo/Claypool 2001 [42]).

Die Diskussion im Seminar hat jedoch auch gezeigt, dass zwar einige Seminarteilnehmer die Chancen dieser Video-Online Stores als gut einstufen, wären aber auf Grund der Qualität und des Preisen noch nicht bereit, diese Art von Streaming zu benutzen. Ein grosser Teil der Teilnehmer räumten diesen Projekten wenig bis kaum Chancen ein, und begründeten ihr Anliegen mit einem scheinbaren rechtlichen Rückschritt. Tauschbörsen im Internet sind dermassen weit fortgeschritten, dass das Internet und sein gesamter Inhalt als inferiores (kostenloses) Gut betrachtet werden, und es alleine aus rechtlicher und pragmatischer Hinsicht extrem schwierig sein wird, diesen Vorsprung wieder einzuholen. An diesem Punkt kommt das Digital Rights Managment (DRM) zur Sprache, da die kommerzielle Nutzung von digitalen Medien zwingend eine wirksame Form der Rechteverwaltung benötigt. Die digitale Form der Inhalte im Internet machen diese sehr anpassungsfähig, kopierbar und einfach zu verbreiten. Die Sicherstellung der Urheberrechte ist eine entscheidende, wenn nicht zwingende Voraussetzung, damit digitale Medien im Internet kommerziell verbreitet werden können. Die gesamte Netzidee basiert auf diesem einfachen Prinzip, was zwingend auch eine Prämisse der Netzökonomie implementiert. Um dies in Zukunft zu verhindern bzw. zumindest zu erschweren, haben die Hersteller von Medienserver- und Playersoftware entsprechende Funktionen zum sog. Digital Rights Management (DRM) in ihre Produkte eingebaut. Diese erlauben die genaue Kontrolle der Nutzung der Medieninhalte. So lässt sich beispielsweise definieren, dass eine Audiodatei nur drei Mal auf eine CD gebrannt werden darf und dass sie nur auf dem PC wiedergegeben werden kann, auf dem

sie ursprünglich vom Medienserver heruntergeladen wurde.

Das internationale sowie technische Fundament des Internets kommt jedoch sehr einer technischen Spielwiese für alle Formen der Ökonomie nahe und bis dato sind gewisse rechtliche Grundlagen noch nicht implementiert worden, wonach teilweise nicht geklärt ist ob nationales oder internationales Recht geltend gemacht werden kann. Alle Formen von solchen Barrieren können vor allem Rund um das Internet umgangen werden und so bleibt die Frage nach dem Urheberrecht von digitalen Inhalten noch ein weites Stück unbeantwortet.

Literaturverzeichnis

[1] Motion Picture Experts Group (MPEG): Expertengruppe für Standards zur komprimierten Übertragung von Audio- und Videodaten. http://www.chiariglione.org/mpeg/

[2] BAKOM, Fernmeldestatistik 2003, http://www.bakom.ch (18.11.2004)

[3] Christensen, Clayton M.: The Innovators Dilemma. When new technologies cause great firms to fail, Boston, 1997

[4] Darwin: The Helix DNA Server, https://helixcommunity.org/2002/intro/server

[5] Dosi, Giovanni: Technological paradigms and technological trajectories, A suggested interpretation of the determinants of technical changes, In: Research Policy 11. Cheltenham, 1982

[6] Dosi, Giovanni: Innovation, Organization und Economic Dynamics, Cheltenham, 2000

[7] Habermas, Jürgen: Theorie des kommunikativen Handelns, Frankfurt am Main, 1981

[8] Hagen, Lutz M.: Riepl's Gesetz im Online-Zeitalter. Eine Sekundäranalyse über Grenzen der Substitution von Massenmedien durch Internet, http://www.gor.de/tband99/pdfs/a_h/hagen.pdf (11.10.2004)

[9] Hartmann, Frank: Technikтheorien der Medien. In: Weber, Stefan (Hrsg.): Theorien der Medien, Konstanz, 2003

[10] Helix, Apple Darwin, http://developer.apple.com/darwin/

[11] EarthCam: Earthcam.com (Verzeichnis vom Internet-Livekameras), http://www.earthcam.com/

[12] Hennen, Leonhard: Technisierung des Alltags. Studien zur Sozialwissenschaft, Opladen, 1992

[13] Hotz-Hart, Beat/Reuter, Andreas/Vock, Patrick: Innovationen: Wirtschaft und Politik im globalen Wettbewerb, Bern, 2001

[14] IETF: The Internet Engineering Task Force, http://www.ietf.org/

[15] RFC 791: Internet Protocol (IP), http://www.ietf.org/rfc/rfc791.txt

[16] RFC 2616: Hypertext Transport Protocol (HTTP), http://www.ietf.org/rfc/rfc2616.txt

[17] Innis, Harold: The bias of communication, Toronto, 2003

[18] IANA: Internet Assigned Numbers Authority, http://www.iana.org

[19] ISO 8601: International Standard Date and Time Notation, http://www.iso.org

[20] JMF: Java Media Framework, http://java.sun.com/products/java-media/jmf/index.jsp

[21] Kiefer, Marie-Luise, Medienökonomie und Medientechnik. In: Altmeppen, Klaus-Dieter/Karmasin, Matthias (Hrsg.): Medien und Ökonomie, Band 1, Wiesbaden, 2003

[22] Latzer, Michael: Mediamatik - Die Konvergenz von Telekommunikation, Computer und Rundfunk, Wiesbaden, 1997

[23] Live.com Streaming Media, http://live.sourceforge.net/

[24] Marr, Mirko: Soziale Differenzen im Zugang und in der Nutzung des Internets. In: Medienheft 11, H. 19.

[25] McLuhan, Marshall H.: Die magischen Kanäle. Understanding Media, Düsseldorf, 1992

[26] Postman, Neil: Das Technopol. Die Macht der Technologien und die Entmündigung der Gesellschaft, Frankfurt am Main, 1992

[27] Real System G2, http://service.real.com/help/library/guides/production/realpgd.htm

[28] RFC 1889: Realtime Transport Protocol (RTP) und Realtime Transport Control Protocol (RTCP), http://www.ietf.org/rfc/rfc1889.txt

[29] RFC 2326: Realtime Streaming Protocol (RTSP), http://www.ietf.org/rfc/rfc2326.txt

[30] RFC 2205: Resource Reservation Protocol (RSVP), http://www.ietf.org/rfc/rfc2205.txt

[31] Rogers, Everett M.: Diffusion of Innovations, New York, 1995

[32] RFC 1980: RTP-Profil für Audio- und Videokonferenzen, http://www.faqs.org/rfcs/rfc1890.html

[33] Seeger, Peter: Technisierung der Medien und ökonomischer Strukturwandel. In: Knoche, Manfred/Siegert, Gabriele (Hrsg.). Strukturwandel der Medienwirtschaft im Zeitalter digitaler Kommunikation, München, 1999

[34] RFC 2327: Session Description Protocol (SDP), http://www.ietf.org/rfc/rfc2327.txt

[35] Synchronisation Multimedia Integration Language (SMIL), http://www.w3.org/AudioVideo/

[36] RFC 793 Transmission Control Protocol (TCP), http://www.ietf.org/rfc/rfc793.txt

[37] Tanenbaum, Andrew S.: Computernetzwerke, 4. Auflage, Pearson-Studium, 2003

[38] Trepte, Sabine/Baumann, Eva/Borges, Kai: Big Brother In: Media Perspektiven 38, H. 12.

[39] RFC 768: User Datagram Protocol (UDP), http://www.ietf.org/rfc/rfc768.txt

[40] RFC 2396: Uniform Resource Identifiers (URI), http://www.ietf.org/rfc/rfc2396.txt

[41] W3C: World Wide Web Consortium, http://www.w3.org/

[42] Wang, Yubing/Zuo, Zheng/Claypool, Mark: An Empirical Study of RealVideo Performance Across the Internet, Hopkinton (MA), 2001

[43] WEMF: AG für Werbemedienforschung, MA-Net 2004, http://www.wemf.ch/d/studien/manet.sthtml (18.11.2004)

[44] Microsoft Windows 2003 Server, http://www.microsoft.com/windowsserver2003/

[45] Zerdick, Axel et al.: Die Internet-Ökonomie. Strategien für die digitale Wirtschaft, Berlin, 2001

[46] Zerdick, Axel et al: E-Merging Media. Kommunikation und Medienwirtschaft der Zukunft, Berlin, 2004

[47] Cable News Network (CNN), http://www.cnn.com

[48] Österreichischer Rundfunk (ORF), http://tv.orf.at

[49] ProSieben, http://www.pro7.de

[50] RTL, http://www.rtl.de

[51] Sat1, http://www.sat1.de

[52] Schweizer Fernsehen DRS, http://www.sfdrs.ch/system/frames/home/videos/index.php

[53] VIVA, http://www.vivamediaag.com

[54] wwiTV, http://mediahopper.com/portal.htm

[55] Zweites Deutsches Fernsehen (ZDF), http://www.zdf.de

Chapter 2

IP Convergent networks

Yves Roesti, Marco Ghinolfi

In this Paper the general issue of IP convergence is addressed. It is shown how the integration of voice, video and data is achieved and why it is beneficial to business. After portraying how IP convergence is affecting the business and its applications, the focus is changed to an individual level, when unified messaging is introduced. Voice over IP (VoIP) and its possible business applications is a famous driver of today's IP convergence. That is why this paper addresses VoIP thoroughly in the second part. After a brief historical introduction, the details of available codecs and their implications on the delay of VoIP System are shown. Those delay issues lead over to the consideration of general Quality of Service concerns. Here, the generic problem of Quality is illustrated and it is shown why the Internet as an IP Network for VoIP transportation is sometimes inapplicable. In a further step we introduce in depth various signaling protocols (SIP, H.323, MGCP) which constitute the basis of VoIP technology. Another chapter is dedicated to the security and privacy issues of VoIP Systems. Numerous forms of attacks and their appropriate counter measures are demonstrated. To enrich this paper with actual VoIP applications, the last part of this paper confers different VoIP software clients, namely Internet Phone and Skype.

Contents

2.1 IP Convergence

In the past, various communication media, such as voice, video and data, have been running in disparate networks, each with its own dedicated physical carrier. In fact they were not only operated in isolation from each other, but also managed by completely separate teams in a company. This parallelism is extremely inefficient and since the base of all media is data (of different kind), it makes sense to consolidate those many media connections to one single network to carry voice, video and data. Thus, it is widely accepted and acknowledged by the communication industry and industry analysts, that the Internet Protocol (IP) will become the universal transport of the future. This way of thinking has drastically increased the significance of the Internet Protocol - or put in words of a Cisco System Manager: "All networks will eventually converge to IP". The change in infrastructure architecture will bring loads of advantages to the business and also to the customer. Since voice, video and data runs through the same medium, they can be made interoperable via specific applications and add additional value.

Some of the key values of this new (IP) paradigm are [22]:

- Enhanced productivity and profitability through new IP-based applications such as integrated multimedia queuing
- Enterprise-wide contact management based on a single set of business rules and supported by normalized consolidated reporting
- Increased customer satisfaction through personalized customer interaction
- Geographic independence of both agent resources and IP-based application servers through the ubiquity of IP transport
- Carrier-quality fault tolerance and system reliability
- Near infinite solution scalability from single-site to multi-site, from network service to provider services
- Lower total cost of ownership, lower capital-equipment investment, single network, and single support staff eliminating the overhead of multiple diverse data, voice, and video networks

Figure 2.1 depicts those benefits for the business using IP convergence. As a result, in near future there will be 2 kinds of costumers:

Customer "Old World" will still be making contact with the company in a traditional fashion, that is via PSTN (Public Switched Telephone Network). To preserve the convergence of the network of the company, his call will not be switched through as it used to be in the past. Instead his call will be transformed into IP packets through a gateway and fed into the company's network.

Customer "New World" has already adopted the new way of communication and is using IP convergence himself (he only has one carrier, which he uses for e-mail, voice, web, etc.). His setup allows a much closer interaction with the company, because he is already using the channels provided.

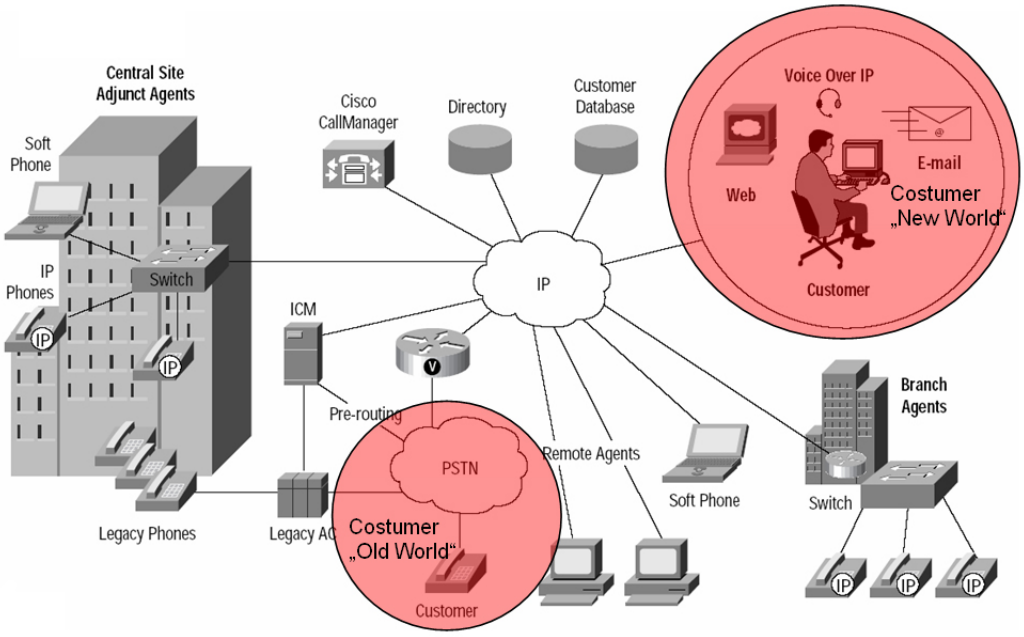

Figure 2.1: The new world of IP convergent networks [22]

2.1.1 Unified Messaging

On an individual level, the IP convergence paradigm offers a tremendous amount of flexibility in working with multi-channel media. This Interoperability between the media can be exploited by an integrated application suite, where you have one application handling all the media. The data collected (be it voice, video or e-mail) can then be viewed on the device suitable to the user's circumstances. For example, while traveling the user won't be urged to have his notebook running to get his e-mails as soon as he exits the plane. Instead he can choose the cell phone to be his medium for delivering the e-mails. This works because the application behind bundles all user data and offers new functionality, like the mentioned e-mail-to-speech module or vice versa (where a voice mail could be transferred to an e-mail text or fax)[22]. Also, it will be possible to establish routing rules for calls or any other incoming stream to fit the user's situation. For example, the "handling application" could be linked to the user's calendar and route calls accordingly. This has a great deal of opportunities, since the call gets somewhat intelligent. If the user sits in a meeting the call can either be routed to his secretary, or be handed over and translated into instant messaging, thus making it possible to handle request even if you're attending a meeting (if this behavior is wanted). Another powerful demonstration of new capabilities is the integration of call related information into the company's (or private) customer relationship management (CRM) database. It allows prioritization, specific routing and also discrimination of costumers (e.g. exploiting the cost of opportunity in a call center application), according to their location. These issues are interesting topics, but can not be discussed in depth here. Concluding, we could state that there is a decoupling between the message and its viewing interface (device). The behavior of arriving messages can be steered by powerful applications opening new ways of flexibility to the user and company.

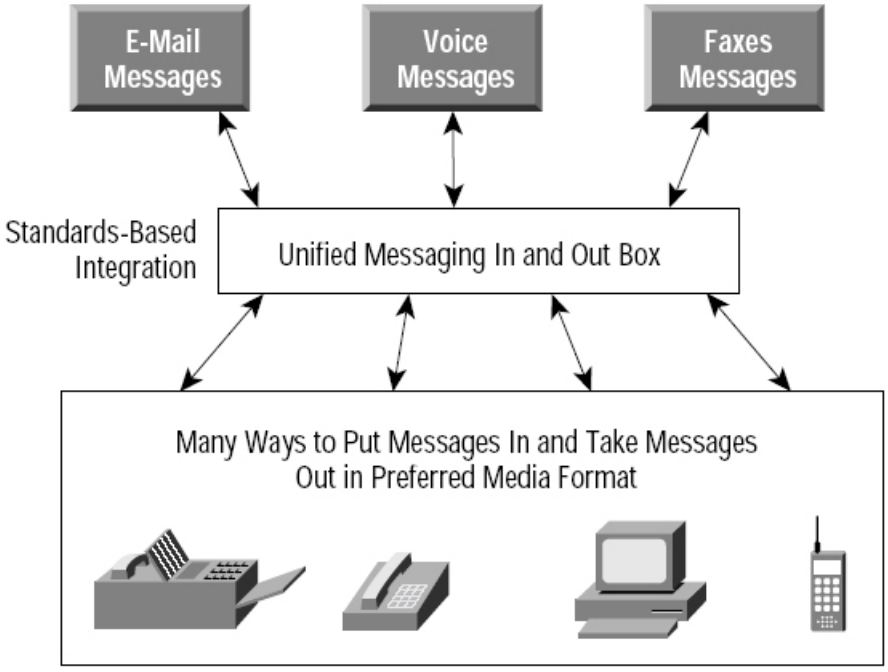

Figure 2.2: The components of unified messaging [22]

2.1.2 Open Systems Architecture behind IP convergence

The Open System Architecture states that innovators of network infrastructure and its software agree that every attempt is made that gateways, applications and clients are made to seamlessly integrate with third party products [22]. For example, the signaling protocols standards (or de facto standards) for voice over IP (VoIP), H.323, MGCP or SIP (please refer to VoIP section for details) will be supported by the underlying network devices. Furthermore, for management of the network devices (Call managers, Customer databases) there is the Telephony Application Programmable Interface (TAPI) and its Java Implementation (JTAPI). Those are protocols that regulate the communication with the layers below.

This is a rough overview of the ambitions for convergence to one medium, the IP network. In the following comments we will concentrate on a specific medium, which has helped to make this IP convergence really a dominant one - the voice over IP (VoIP).

2.2 VoIP Technology

2.2.1 History

Alexander Graham Bell and Elisha Gray both invented the telephone at about the same time in 1876. At the beginning telephones were not much more than entertaining. But

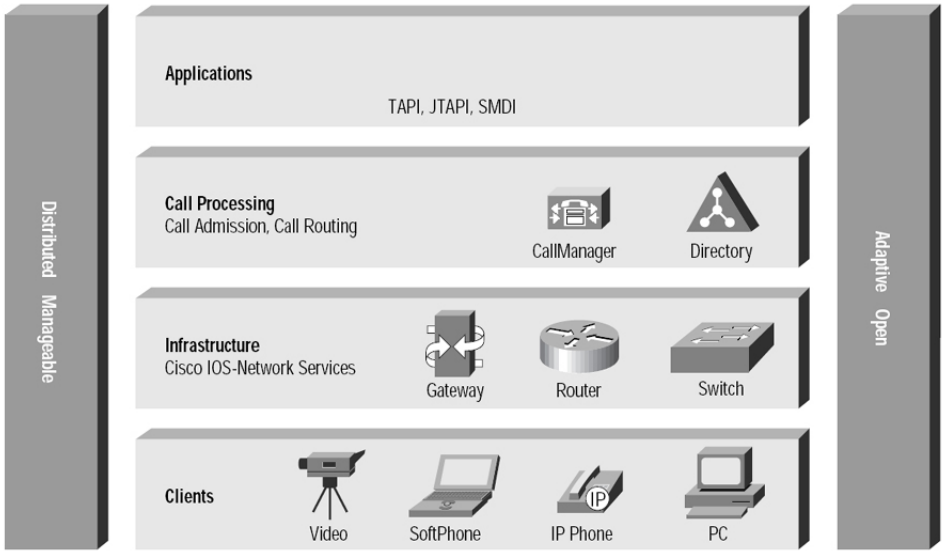

Figure 2.3: CISCO's AVVID (Architecture for Voice, Video and integrated Data [22]

newspapers and banks soon realized the advantages of the telephone to exchange information very quickly. In 1891 direct dialing was invented so there was less need for operators to connect the calls. Great innovations were made during the both world wars with the help of heavy spending by the defense departments. In the following years the number of telephone subscribers grew so big that alpha-numeric numbers were no longer possible. Only numeric codes could be used from this point in time. In the 1960s the telephone user dialed the number with a rotating disk on the telephone. The number of impulses resulting when the dial rotated back indicated the digit that was dialed. These pulses where identified by the local exchange.

Many of today's phone users think that this is a disturbing and obsolete way to dial a number. The fact is a considerable number of these older units are still in operation in many countries all over the world. Modems and Telefaxes started to be used in the late 1960s. Bit rates were between 300bps and 600bps. Thirty years ago, ISDN was going to be the answer to all of our problems, using circuit-switching for both voice and data [13]. Fiber optics cables were first used in 1977. Fiber optic cable was the preferred method of telephone transmission, since it could carry a much higher volume of calls with much less interference. Voice over Internet Protocol (VoIP) soon became popular for telephone communications because it avoids the toll charges of standard telephone connections. With broadband connections the increased data throughput enabled businesses to use VoIP in conjunction with other Internet services like data sharing and video conferencing. With the money saved using VoIP, it seems obvious that using analog phone lines for telephone conferencing will soon be a thing of the past [14].

VoIP came into existence as a result of work done by a few hobbyists in Israel in 1995 when only PC-to-PC communication was in vogue. Later on during 1995, Vocaltec, Inc. released Internet Phone Software. The software was used to compress the voice signal, convert it into voice packets, and then finally ship it out over the Internet. This particular technology worked as long as both the caller and the receiver had the same tools and software. However, the sound quality was not even close to that of the standard

equipment in use at that point in time. This attempt can be termed as the first IP phone that came into existence [15]. In 1998 several companies set up gateway to allow PC-to-Phone and later even Phone-to-Phone connections. Of course these services still require a PC to originate the call even if it's a Phone-to-Phone connection. In the year 2000 VoIP was 3% of the voice traffic in North America and it's expected to grow to 25 - 40%. So IP convergent networks grow more important as they go along. Services like video telephony and distributed working are emerging technologies that are about to come on line.

2.2.2 PSTN (Public Switched Telephone Network)

The traditional telephone system which is based on copper wires carrying analog voice data is called PSTN. This stands for Public Switched Telephone Network. It is the world's collection of interconnected voice oriented public telephone networks, both commercial and government-owned. The PSTN is still the most used telephone network. The PSTN's primary characteristics are:

- analog access, 300-3,400 Hz
- circuit-switched duplex connection
- switched bandwidth, 64 kbit/s, or 300-3,400 Hz for analog exchanges
- immobility or, at best, very limited mobility; and
- many functions in common with another bearer network: N-ISDN

Telephone service carried by the PSTN is often called Plain Old Telephone Service (POTS) [18].

2.2.3 VoIP Codec

Transporting IP packets with media data (voice, video or other real time data) through any kind of network, especially the Internet takes a certain time (also called transportation time). This fact makes it necessary to apply a certain level of compression to those packets in order to keep the latency down. There is always a trade off between data transferred and voice quality. Bandwidth and routing protocols (e.g. RSVP, please refer to QoS) play an important role within this trade off - they determine how fast a VoIP packet is routed forward along the communication path and therefore how good the quality can be. In the end, the codec used (in other words the compression applied) must fit those given conditions.

A high compression isn't always applicable if the bandwidth is tight - indeed it downsizes the amount of data sent, but also increases the processing power needed to execute the compression/decompression algorithms duly. An additional handicap when using the Internet as a medium is the drop of packets, which has to be compensated by the codec.

Table 2.1: Overview of Codecs [31]

Codec	Description	Transfer rate Payload/w. Header	Delay	Voice quality	MOS
G.711	Pulse Code Modulation (PCM)	64 kbit/s / 80 kBit/s	0,75ms	ISDN	4.4
G.726	Adaptive Differential Pulse Code Modulation (ADPCM)	16-40 kbit/s / 40-60 kbits/s	1 ms	Mobile Phone	3.9
G.728	Low Delay Code Excited Linear Prediction (LD-CELP)	16 kbit/s / 32 kbits/s	3-5 ms	Almost ISDN	3.7
G.729/ G.729A	Conjugate Structure Algebraic Code Excited Linear Prediction (CSACELP)	8 kbit/s / 31,2 kbit/s	10-20 ms	Better than G.723.1	3.4
G.723.1	Multiple Maximum Likelihood Quantization (MPMLQ)	6,3 kbit/s	30-60 ms	Good	3.9
G.723	Algebraic Code Excited Linear Prediction (ACELP)	5,4 kbit/s / 21,9 kbit/s	40-60 ms	Good	3.5

Commonly this drop averages around 5% of all the packages sent. The codec has to provide a correction algorithm to even out the lost packages and also sort the received packages in the right order. Techniques used to fulfill this challenge are called "Forward Error Correction" and "Jitter Buffering" (See details below). Choosing the right codec is, according to the implementations above, a consideration and balance of all parameters mentioned and therefore a very situational decision.

Each codec provides a certain quality of speech. The quality of transmitted speech is a subjective response of the listener. A common benchmark used to determine the quality of sound produced by specific codecs is the mean opinion score (MOS). With MOS, a wide range of listeners judge the quality of a voice sample (corresponding to a particular codec) on a scale of 1 (bad) to 5 (excellent). The scores are averaged to provide the MOS for that sample. The following table 2.1shows the relationship between codecs and MOS scores [1].

Typically, for VoIP applications the Codec G.729A is used. Considering the IP overhead and the compression, a bandwidth of 10 kbits/s thus is required for voice communication. This calculates to 1.25 kbytes/s, which has to be supported through the whole network [9].

Although it might seem logical from a financial standpoint to convert all calls to low-bit rate codecs to save on infrastructure costs, additional care should be exercised when designing voice networks with low-bit rate compression. There are drawbacks to compressing

voice. One of the main drawbacks is signal distortion due to multiple encodings (called tandem encodings). For example, when a G.729 voice signal is tandem encoded three times, the MOS score drops from 3.92 (good) to 2.68 (unacceptable). Another drawback is codec-induced delay with low bit-rate codecs. The issue of codec induced delay and general delay concerns will be discussed in the next section.

2.2.4 Delays in VoIP Systems

Compression takes time, which adds to the delay. Of course a higher compression is much more subjected to delay (compare to codec table above) then lower compression. The delay is not only caused by the codec, but also from the transportation distance of the IP-packet. Figure 2.4 shows the time consumption using the G.729 compression algorithm. According to the 2.2, the basic block size of G.729 is 10-20 ms (Here: Assumed 10 ms). If you would use the G.723.1 instead of the G.729, a block size of 30 ms would result (which is three time the delay).

Note: The blocks shown correspond to PCM (Pulse Code Modulation - a technique used to digitalize voice fragments) blocks.

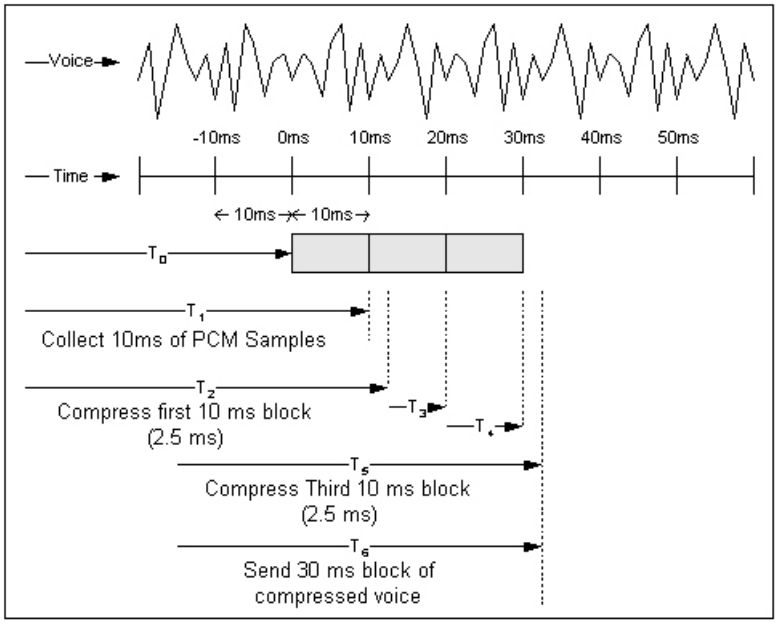

Figure 2.4: Example of PCM block compression [23]

The PCM blocks are being collected for 10 ms and then compressed by the algorithm (Codec delay). If this is done three times, a Packetization delay of 30 ms results. Packetization delay may also be called Accumulation delay, as the voice samples are accumulated in a buffer before being released [2]. In order to measure the transportation delay, apart from the discussed Packetization and Codec delay, you will have to look at the underlying network. Here, the network switching delay is prominent, because the routing and queuing algorithm in every passed node of the network adds up to the delay. Hence, those

Table 2.2: Overall delay specifications [32]

Range in Milliseconds	Description
0-150	Good for most user applications.
150-400	AAcceptable, starting from 250 ms the voice quality is influenced negatively.
Above 400	Unacceptable for general VoIP purposes. However, it is recognized that in some exceptional cases this limit will be exceeded.

network switching delays are the hardest to estimate, because they are composed of many factors.

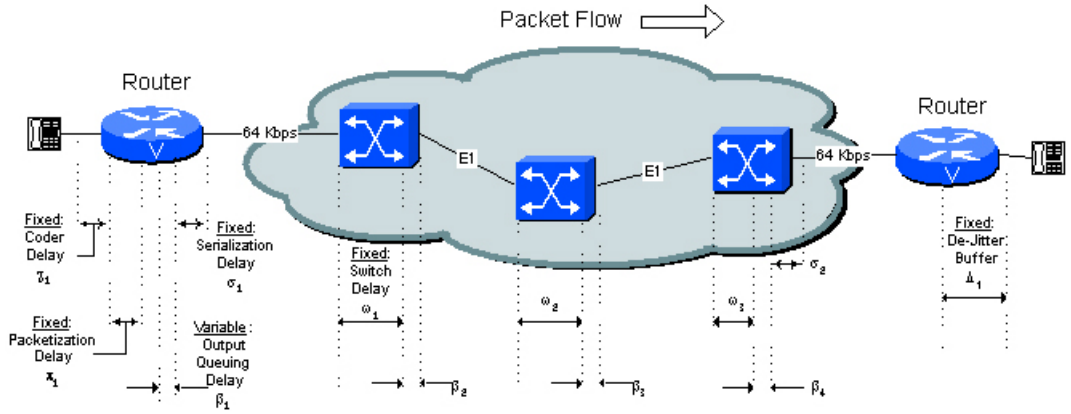

Figure 2.5: Delay sources [24]

Finally, once the voice packet has reached the end of the line there is the De-jitter Delay, which tries to even out the variable arrival rate of the packets over a buffer. Packets can arrive too early, doubled, incorrect or even too late. Up to 5% of mislead or lost voice data can be managed by the buffer. Proper handling of the de-jitter buffer is critical. If samples are held for too short a time, variations in delay may cause the buffer to underrun and cause gaps in the speech. If the sample is held for too long a time, the buffer can overrun, and the dropped packets again cause gaps in the speech. Lastly, if packets are held for too long a time, the overall delay on the connection may rise to unacceptable levels. All those delays add up to an overall delay figure. The following table provides the level of delays specified for a network provider. If the delay rises over the critical mark of 400 ms, the voice is severely delayed; this can be compared to a cellular phone session with your voice partner having bad reception [2][9]. The next chapter picks up the delay issue again and introduces a general view of quality of service (QoS) in VoIP Systems by looking at other drawbacks.

2.2.5 Quality of Service (QoS)

Because IP does not by default provide any mechanism to ensure that data packets are delivered in sequential order, or provide any Quality of Service guarantees, implementations

of VoIP face problems dealing with latency and possible data integrity problems. One of the central challenges for VoIP implementers is restructuring streams of received IP packets, which can come in any order and have packets missing, to ensure that the ensuing audio stream maintains a proper time consistency. Another important challenge is keeping packet latency down to acceptable levels so that users do not experience significant lag time in the conversation.

Solutions to these problems:

- Certain hardware solutions can distinguish VoIP packets and provide priority queuing for this class of service (CoS).
- Alternatively packets can be buffered but this can lead to an overall delay similar to that encountered on satellite circuits.
- The network operator can also ensure that there is enough bandwidth end-to-end to guarantee low-latency low-loss traffic: this is easy to do in private networks, but much harder to do in the public Internet.

Normal runtime for traditional circuit switching technology: 90-250ms (depending on the route). This provides a good quality of conversation (see section "Delay in VoIP Systems"). Consequently for data networks that want to incorporate voice service, those runtimes have to be kept as low as possible. As mentioned above, the specific hardware equipment used in networks have to be able to identify VoIP Packets and give them priority for routing. To establish this sort of guarantee for fast processing in order to lower the runtime, certain Quality of Service (QoS) and Class of Service (CoS) mechanisms have to be deployed [3].

On one hand the runtime can be reduced by adjusting the (that means applying the right) codec used to compress the voice or video data (please refer to the VoIP codec section for more information). This is an isolated measurement on the data packets. On the other hand there are approaches on the network that carry those packets. The Resource Reservation Protocol (RSVP) is a protocol that allows a client to indicate its required bandwidth and the maximum delay time for its intended service (be it voice or video) to the network components. A request is passed along all possible communication paths and the delays and bandwidth constraints of each device are added up, resulting either in a declining or definite assertion for the requirements of a communication path posed initially by the client [7].

The third bullet point in the listing above addresses clearly the Internet as a restraint for QoS guaranty. Because of its heterogeneous nature and its many mediators (meaning server, router) the necessary enforcement for adaption of protocols enabling Qos/CoS is simply impossible. The RSVP protocol may fit perfect as an enabler for VoIP in a small network where all the communication devices go by proposed standards, but in the wide internet such protocols are rarely implemented. This circumstance explains why VoIP is widely popular in inter-office application, where the infrastructure is laid out to cope with the necessary QoS. It also makes clear why VoIP using the Internet is still a utopia.

This fact is interesting because even the nowadays adopted internet protocol itself (ipv4) actually has an attribute for priority in packets.

Consequently everything goes by ”Best Effort” in the Internet, which is a killer criterion for time critical applications like VoIP. Some believe that the successor of the ipv4, the ipv6 will achieve true QoS with its setup, but in reality everything is dependent on the infrastructure of the provider and their will to adapt higher worthy IP streams. Having defined what quality is and how it can be achieved in VoIP systems, we will now present how sessions are established. The crucial VoIP signaling protocols are the topic in the next section.

2.2.6 VoIP protocols

In most implementations, the RTP protocol is used to transmit VoIP traffic (or general media). As shown in Figure 2.6 those implementations make use of the existing internet protocol stack up to the TCP/UDP Layer. RTP defines a standardized packet format for delivering audio and video over the Internet. It is frequently used in streaming media systems (in conjunction with RTSP) as well as videoconferencing and push-to-talk systems (in conjunction with H.323 or SIP), making it the technical foundation of the Voice over IP industry. Specifically, RTP provides the transport to carry the audio/media portion of VoIP communication. RTP is used by all the VoIP signaling protocols.

H.323

Aboriginal H.323 was designed for video communication over TCP/IP (Transmission Control Protocol / Internet Protocol). It was normalized by ITU-T (International Telecommunication Union) in 1996. For H.323 a whole protocol family was developed (Figure 2.6). There is the H.225.0 for setup, the Q.931 for signalling H.245 for telephony and H.450 for further services.

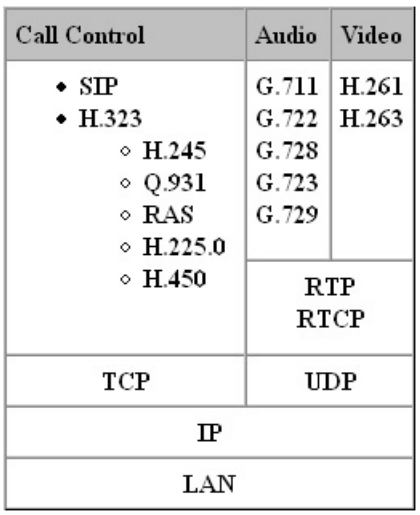

Figure 2.6: A possible protocol stack for VoIP applications [25]

H.323 also defines also gateway and gatekeeper functions. The gateway allows transitions to other voice nets, the gatekeeper manages bandwidth and translation of the symbolic address to the IP address. As soon as the connection is stable H.245 carries out its handshake operation. It now checks if codecs for voice or video compression are available. Standard is the G.711 codec for voice. To phone out of a LAN to the PSTN there is need of a VoIP - gateway (see Figure 2.7 and 2.8). It translates the data volume from the G.723.1 codec to the G.711 codec. If there is no need to connect to the PSTN, there is also no need for a gateway. Core of the PSTN/IP - gateway is the gatekeeper. It serves the emulation of the setup of the connection to the PSTN, the translation of PSTN data to IP packages and back and the translation of the IP address to the PSTN telephone number and back.

Figure 2.7: Telephony from LAN to PSTN [26]

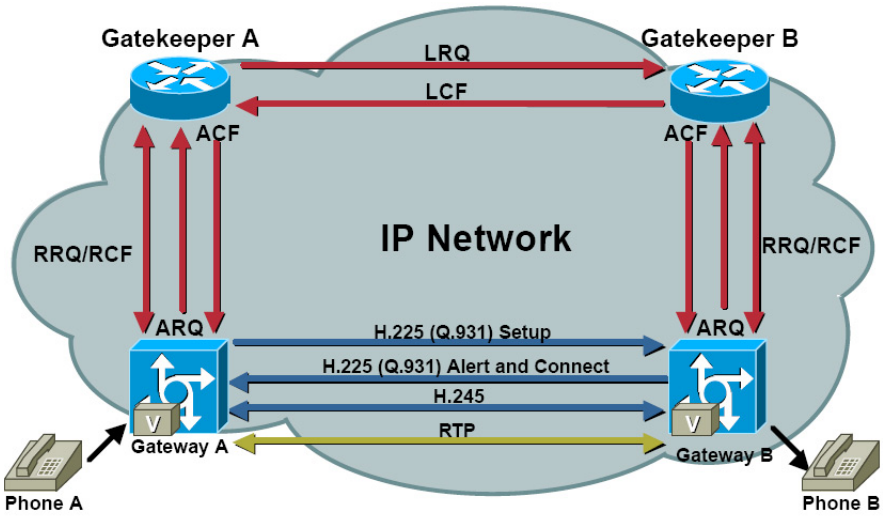

Figure 2.8: Gateways and their gatekeepers [27]

SIP

SIP is the abbreviation for Session Initiation Protocol. It was normalized by IETF (Internet Engineering Task Force) in 1999. It was developed to connect users to conferences. It is the new challenger to the H.323 protocol. SIP is based on Internet protocols like SMTP (Simple Mail Transfer Protocol) or HTTP (Hypertext Transfer Protocol). So SIP is very easy to use and easy to implement in applications and devices. Information are transmitted in very simple plain text, unfortunately it's easy to manipulate. SIP is used as signal control protocol for the setup, modifications and terminating of sessions between one or multiple users in an IP based network. Sessions can be voice calls over IP network most likely the internet, internet multimedia conferences or multimedia transfers. Session

Table 2.3: SIP requests [33]

Invite	The other side is invited to a session. This is like the signalling of a ringing telephone.
Ack(knowledge)	With this request a connection is acknowledged.
Bye	This request is executed when one of the participants terminates the connection.
Cancel	This request is executed when the connection is terminated after a given time.
Options	This request is for transmitting user data.
Register	This request is for registering the location of the client so that a call can be sent to the client.

participants are connected through a net of unicast connections or multicast connections or a combination of both. SIP has several dialogs to set up a session between two participants. Dialogs are made up of a request and a response. SIP knows following requests [34]:

A SIP-based network is built with several components (Figure 2.9):

- SIP User Agent-any network end-point that can originate or terminate a SIP session. This might include a SIP-enabled telephone, a SIP PC-client (known as a ”softphone”), or a SIP-enabled gateway.
- SIP Proxy Server-a call-control device that provides many services such as routing of SIP messages between SIP user agents
- SIP Redirect Server-a call-control device that provides routing information to user agents when requested, giving the user agent an alternate uniform resource identifier (URI) or destination user agent server (UAS).
- SIP Registrar Server-a device that stores the logical location of user agents within that domain or sub-domain. A SIP registrar server stores the location of user agents and dynamically updates its data via REGISTER messages.

Differences between H.323 and SIP

SIP is very easy to integrate in an existing IP network, SIP uses all the available services like HTTP, SMTP, MIME, URL and DNS. There is no need for a gatekeeper as it exists in H.323. SIP uses a proxy, so there’s no problems with heavy traffic (Figure 2.10).

MGCP

The abbrieviation MGCO stands for Media Gateway Control Protocol. It’s one of the newer technologies and a competitor of SIP. MGCP is a device control protocol developed by IETF and destined to control devices, like Media Gateways, by using text format

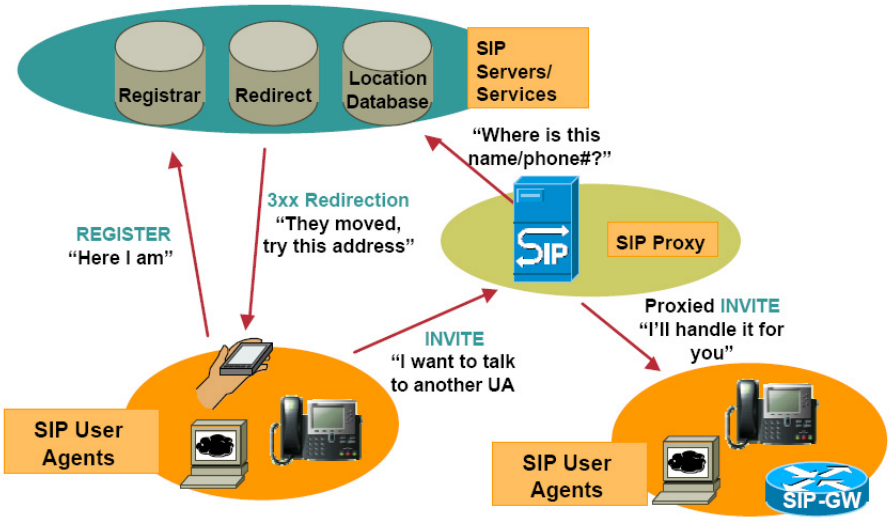

Figure 2.9: SIP network [28]

Table 2.4: H.323 vs. SIP

VoIP standard	H.323	SIP
Philosophy	Precise defined architecture, rules and control	Setup and termination of two or more participants. Only the necessary is defined
Requirements	Telecommunicating technology	IP network (e.g. Internet)
Backward compatibility	Adding of new performance characteristics	Older performance characteristics are replaced by newer ones
Architecture	Control by the server	Control by the client

messages to set up, manage, and terminate multimedia communication sessions in a centralized communications system. The difference between MGCP and other multimedia control protocol systems is that MGCP allows the endpoints in the network to control the communication session. MGCP is a protocol that operates between a Media Gateway (MG) and a Media Gateway Controller (MGC) - (also known as Call Agents or Soft Switches), allowing the Media Gateway Controller to control the Media Gateway. MGCP enforces the Media Gateway as the fundamental component of multipoint, next generation, converged networks. MGCP was developed as part of the convergence movement, which brings voice and data together on the packet-switched Internet.

In a classic call scenario between two endpoints (Figure 2.11), the call agent(s) controlling the endpoints will establish two connections. Each connection will be designated locally by a connection identifier, and will be characterized by connection attributes.

Assume that User A wants to call User B. User A is located within the IP network, served from a residential gateway and User B is located off-net via the PSTN. When User A picks up the phone, a ”notify off hook message” is sent from the residential gateway to the Call Agent. The Call Agent asks the gateway to create a connection on the endpoint that went off hook by sending a create connection command. The gateway acknowledges to the Call agent the create connection command plus provides a session description. The

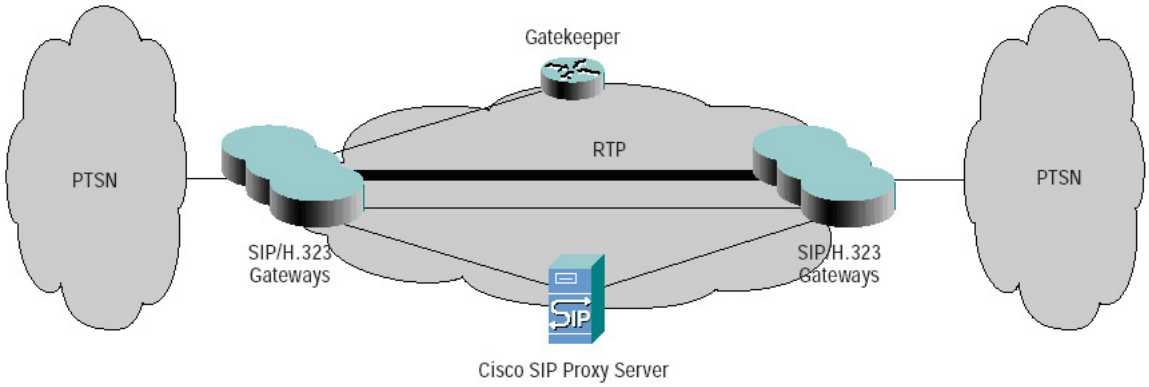

Figure 2.10: SIP and H.323 [29]

session description contains information required by a third party, in this case the trunking gateway, to send packets toward the newly created connection. The ”Session Description Protocol” is used for this and contains such things as User A’s IP address, the UDP port to identify the session, packetization parameters such as compression techniques, and a media type such as RTP audio (voice). The trunking gateway responds to the Call Agent providing its own session description.

The Call Agent uses a modify connection to provide the session description from the trunking gateway to the residential gateway. A two-way full duplex communication can now be set up between the residential gateway and the trunking gateway. When a connection is set up between endpoints, RTP (Real-time Transport Protocol) is used. RTP runs of course on top of UDP because it has multiplexing capabilities, and acknowledgement of packet delivery is not required.

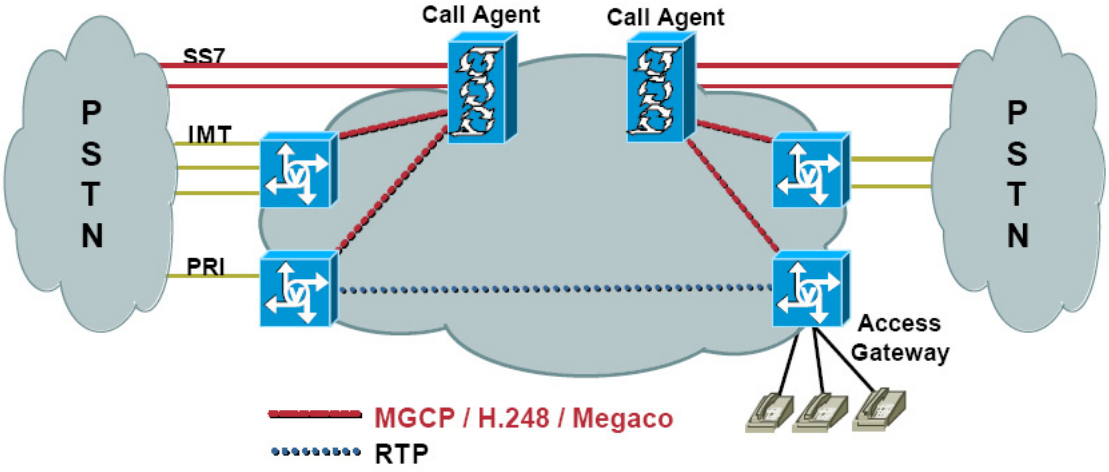

Figure 2.11: Call transport in the MGCP [30]

2.2.7 Security / Privacy

Once voice is converged with data on the network, a company's voice systems are suddenly vulnerable to many of the same kinds of attacks that occur on the data side. Phones can suddenly become destinations for spam. Hackers can target phone systems with denial of service attacks, or program a company's phones to call other businesses, shutting down the second company's phone systems. People can spoof a phone's IP address and make calls that are billed back to the company. And as with a traditional phone system, calls can be intercepted and listened to [12]. Nowadays most systems are based on the Session Initiation Protocol (SIP, see above). And because of this standardization attacks on VoIP Systems could increase.

Sniffing services in an IP convergent network - example

It's relatively easy to "sniff" VoIP calls. What you need is a node through which data is sent. There are two communication channels necessary for correct functioning. One is for the control of the session (signal path) and the other is for the multimedia data (data path). To control a session you need a signal protocol. The signal protocol controls services like calling, terminating, switching and so on and it handles parameters like codecs. Codecs are used to structure and compress the data. The RTP (Real-time Transfer Protocol) / UDP is commonly used to transport media data. With a sniffer like "Aldebaran" you can forward filtered UDP Packages of the data stream to any host you like. On this host machine a tool like Java Media Studio (JMS) should be installed to play the data stream. JMS allows you to open RTP-Sessions. It also contains a lot of codecs to play the data. In a LAN attackers don't even need to capture a nod where the data is sent through. Small programs can confuse the switches in a way that they are sending data to every port. So the data can be sniffed quite easily. Not only voice data is easy to sniff. Video data and corporate distributed working can also be copied and redirected to a host that recodes the data. So the IP convergence let's you easily spy on various data, no matter if it's video or voice.

Attacks

The main security concern nowadays is the denial of service attack (DoS attack) or the distributed denial of service attack (DDoS attack). Almost anyone with access to a network, that's used for IP convergent services, is potentially able to knock the server out of service. It is, compared to alternate services, more or less trivial to completely disable a VoIP phone by sending malformed signaling messages or simply overwhelming it with junk. Vendors haven't addressed this problem seriously yet but it probably will be necessary to find a solution. Otherwise vendors will lose costumers.

Another form of attack is the so called "man-in-the-middle attack". A man in the middle attack (MITM) is an attack in which an attacker is able to read, insert, and modify at will messages between two parties without either party knowing that the link between them has been compromised. The attacker must be able to observe and intercept messages

going between the two victims. Suppose Angela wishes to communicate with Ben, and that Marco wishes to eavesdrop on the conversation, or possibly deliver a false message to Ben. To get started, Angela must ask Ben for his public key. If Ben sends his public key to Angela, but Marco is able to intercept it, a man in the middle attack can begin. Marco can simply send Angela a public key for which he has the private, matching, key. Angela, believing this public key to be Ben's, then encrypts her message with Marco's key and sends the encyphered message back to Ben. Marco again intercepts, decyphers the message, keeps a copy, and reencyphers it (after alteration if desired) using the public key Ben originally sent to Angela. When Ben receives the newly encyphered message, he will believe it came from Angela. Kevin Kealy, a scientist who works for AT&T explained that it is possible to inject swear words into a conversation of two subscribers of a VoIP service so that even the speaker can't hear the words but the hearer can. Voiceprints of such examples recorded by the FBI were so good that FBI rated them as "genuine" [16].

Another threat that could be a real concern is spam. Spam in E-mail transfer is a very disturbing thing and there is no 100% protection against it. Spam could come to services as VoIP and relating technologies. Automated ad calls that fill your voice box could be more than just possible. The term used to describe such spam is SPIT (Spam over IP Telephony).

Protection

IPsec is one technology that could help to protect the privacy of services such as VoIP. IPsec stands for Internet Protocol Security and it is obligatory in IPv6. In IPv4 it is optional and not as widely used as it could be. IPsec operates at layer 3 of the OSI model, so it is suitable for both TCP and UDP protocol. For comparison: The SSL protocol cannot encrypt UDP packages because SSL is on a higher layer. IPsec is a set of interdependent protocols consisting of a protocol that is for securing packet flows and one for the key exchange. The sending and receiving devices have to share their public keys to work. There are two types of encryption. One is the transport. It encrypts only the payload but leaves the header untouched. The other is the tunnel, where the IPsec encrypts also the header which is more secure [17][18].

Another protection is provided by signal protocols themselves. The Session Initiation Protocol (SIP) provides similar functionality to H.323. SIP is specified by the Internet Engineering Task Force (IETF) for initiating a two-way VoIP communication session. SIP prevents an attacker from eavesdropping on a call. It also prevents someone from hijacking a call but its does not prevent DoS attacks. SIP secures the media in the call. Even if a hacker manages to gather all the packages, he won't be able to decrypt it. This is done using SRTP (Secure Real-time Transfer Protocol), which encrypts and authenticates each media packet. Unfortunately these techniques aren't in widespread usage. In a network where packages are broadcasted and no such protection exists, it's quite easy for a hacker to extract and analyze the data [19].

2.2.8 Software Clients for IP-Telephony

There are quite a few software clients for Internet Telephony. But it started with only one. As mentioned before, Vocaltec released their Internet Phone software in 1995. The quality wasn't as good as you might think. The bandwidth of most internet users wasn't much more than 28800bps. So this software was not a huge seller at that point. As time went by and demand grew bigger several other products were created and brought to the market. One of these products became very popular and it's not the latest version of Vocaltec's Internet Phone. They still survived and sold a lot of copies of Internet Phone which is now released as version 5. The software that is now number one in the USA and probably all over the world is named Skype. The career of one of the co-founders of the Skype Company, Nilkas Zennström, is very interesting. Niklas Zennström started his professional career at Tele2, when there were only 23 employees. Now the company is the leading consumer oriented pan-European telecom operator present in 23 countries. He later co-founded and served as CEO of KaZaA, the world's most downloaded Internet software to date with more than 370 millions downloads [20]. Internet Telephony software nowadays not only allows you to talk over the internet with other persons, who use the same software on a computer on the other side, but gives you the possibility to access the PSTN. First you have to buy credits paying by credit card. After that you can dial telephone numbers in other countries. Costs are around calling costs of a local call in the destination region. This means you pay as much as you would if you actually were in the other country and phoning over the PSTN there. So this would be a PC-Phone connection.

An important difference to the hardware based VoIP is the fact that internet telephony cannot guarantee the same quality of service as the hardware based approach. The main difference lies in the technology. While internet telephony bases on the public internet with no control at all, hardware VoIP providers can control and prioritize calls on their wires. Hardware VoIP providers have a work with a direct broadband line from the customer to their infrastructure (e.g. over the television cable). The call is then forwarded most likely to a PSTN phone. The biggest part of the way is in hand of the provider whereas internet telephony providers as a rule do not provide internet access. So because internet telephony providers have no chance to own the net or at least the backbone of the internet, they cannot guarantee fast and reliable sessions. They depend on the quality of the internet, provided by other infrastructures and companies.

2.3 Conclusion

As we have seen, IP convergence leads to lots of new business models and applications. It will change the way businesses interact with their customers in the future. With its "one line fits all" credo, IP networks are excellently suited to fulfill such strategies. A powerful application is needed though to forward the benefits from the underlying network architecture to the individual. The user himself gets channel independent and this adds a great deal to his productivity. A peculiarity of the IP convergence is voice over IP, which has been discussed in detail. We have shown that VoIP has great market potential, if the

quality issues are addressed properly. This can be done by choosing the right codec and by making specific adaptions to the layout of the network. On top of all, there has to be a protocol that ensures the prioritization of voice packets. In the VoIP signaling protocol section we looked at the procedures to establish and tear down call connections. This activity forms the basis for a successful call. Furthermore we pointed out that, like every other digitalized medium, VoIP calls can be intercepted, disturbed, spamed (spitted) or hijacked. To absorb such threats on privacy various retaliatory actions were introduced. In the end we presented software clients that integrate all the technology that has been reviewed in this paper.

As a last outline, the following numeration tries to summarize the economic benefits from VoIP technology:

- Advantages compare to conventional circuit switching (PSTN): The traditional telephony is using, due to its direct, switched-through connection, full bandwidth, which is a wasteful strategy (in term of bandwidth efficiency). With packet switching, IP packets are only sent when they have to be sent (e.g. when someone speaks)
- Cost savings: By moving voice traffic to IP networks, companies can reduce or eliminate the toll charges associated with transporting calls over the Public Switched Telephone Network (PSTN). Service providers and end users can also conserve bandwidth by investing in additional capacity only when it is needed. This is made possible by the distributed nature of VoIP and by reduced operations costs as companies combine voice and data traffic onto one network.
- Open standards and multivendor interoperability: By adopting open standards, both businesses and service providers can purchase equipment from multiple vendors and eliminate their dependency on proprietary solutions.
- Integrated voice and data networks: By making voice "just another IP application", companies can build truly integrated networks for voice and data. These integrated networks not only provide the quality and reliability of today's PSTN, they also enable companies to quickly and flexibly take advantage of new opportunities within the changing world of communications.

Bibliography

[1] Cisco - Understanding Codecs: Complexity, Hardware Support, MOS, and Negotiation; http://www.cisco.com/warp/public/788/voip/codec_complexity.html, January 2005.

[2] Understanding Delay in Packet Voice Networks [Voice Quality] - Cisco Systems; http://www.cisco.com/en/US/tech/tk652/tk698/technologies_white_paper09186a00800a8993.shtml, January 2005.

[3] Voice over IP - Wikipedia, the free encyclopedia; http://en.wikipedia.org/wiki/Voice_over_IP, January 2005.

[4] Voice over IP - Wikipedia, the free encyclopedia; http://de.wikipedia.org/wiki/Internettelefonie, January 2005.

[5] Voice over IP - Wikipedia, the free encyclopedia; http://en.wikipedia.org/wiki/H.323, January 2005.

[6] Voice over IP - Wikipedia, the free encyclopedia; http://en.wikipedia.org/wiki/Session_Initiation_Protocol, January 2005.

[7] QoS - Quality of Service (Voice over IP); http://www.elektronik-kompendium.de/sites/net/0905131.htm, January 2005.

[8] VoIP - Voice over IP; http://www.elektronik-kompendium.de/sites/net/0503131.htm, January 2005.

[9] Codecs zur Sprachdigitalisierung (Voice over IP); http://www.elektronik-kompendium.de/sites/net/0905121.htm, January 2005.

[10] Understanding H.323 Gatekeepers [IP Telephony/Voice over IP (VoIP)] - Cisco Systems; http://www.cisco.com/en/US/tech/tk652/tk701/technologies_tech_note09186a00800c5e0d.shtml, January 2005.

[11] IEC: H.323; http://www.iec.org/online/tutorials/h323/, January 2005.

[12] The security risks of VoIP; http://searchcio.techtarget.com/originalContent/0,289142,sid19_gci1032194,00.html, January 2005.

[13] http://www.teracomtraining.com/courses/301/teracom-brochure-course-301.pdf, January 2005.

[14] History of the Telephone – from Bell to VoIP and Beyond; http://ezinearticles.com/?History-of-the-Telephone—-from-Bell-to-VoIP-and-Beyond&id=6957, January 2005.

[15] Brief History of VoIP - Intertangent Technology Directory; http://www.intertangent.com/023346/Articles_and_News/1477.html, January 2005.

[16] VoIP security a moving target; http://www.nwfusion.com/news/2004/102504von.html, January 2005.

[17] IPsec - Wikipedia, the free encyclopedia; http://en.wikipedia.org/wiki/IPSEC, January 2005.

[18] What is IPsec? - A Word Definition From the Webopedia Computer Dictionary; http://www.webopedia.com/TERM/I/IPsec.html, January 2005.

[19] VoIP Security Safeguards and SIP; http://www.eweek.com/article2/0,1759,1593993,00.asp, January 2005.

[20] Skype Founders; http://www.skype.com/company/founders.html, January 2005.

[21] Media Gateway Control Protocol (MGCP) Technology; http://www.ixiacom.com/library/technology_guides/tg_display.php?skey=mgcp, January 2005.

[22] Cisco Systems; http://www.cisco.com/warp/public/cc/so/neso/vvda/iptl/avvid_wp.pdf, January 2005.

[23] Understanding Delay in Packet Voice Networks [Voice Quality] - Cisco Systems; http://www.cisco.com/en/US/tech/tk652/tk698/technologies_white_paper09186a00800a8993.shtml, January 2005.

[24] Understanding Delay in Packet Voice Networks [Voice Quality] - Cisco Systems; http://www.cisco.com/en/US/tech/tk652/tk698/technologies_white_paper09186a00800a8993.shtml, January 2005.

[25] Cisco Systems; http://www.cisco.com/warp/public/cc/so/neso/vvda/iptl/avvid_wp.pdf, January 2005.

[26] H.323 (Voice over IP); http://www.elektronik-kompendium.de/sites/net/0905101.htm, January 2005.

[27] Cisco Systems; http://www.cisco.com/warp/public/cc/so/neso/vvda/iptl/avvid_wp.pdf, January 2005.

[28] Media Gateway Control Protocol (MGCP) Technology; http://www.ixiacom.com/library/technology_guides/tg_display.php?skey=mgcp, January 2005.

[29] Cisco Systems; http://www.cisco.com/warp/public/cc/so/neso/vvda/iptl/avvid_wp.pdf, January 2005.

[30] Media Gateway Control Protocol (MGCP) Technology; http://www.ixiacom.com/library/technology_guides/tg_display.php?skey=mgcp, January 2005.

[31] Codecs zur Sprachdigitalisierung (Voice over IP); http://www.elektronik-kompendium.de/sites/net/0905121.htm, January 2005.

[32] Understanding Delay in Packet Voice Networks [Voice Quality] - Cisco Systems; http://www.cisco.com/en/US/tech/tk652/tk698/technologies_white_paper09186a00800a8993.shtml, January 2005.

[33] SIP - Session Initiation Protocol (Voice over IP); http://www.elektronik-kompendium.de/sites/net/0905111.htm, January 2005.

[34] SIP - Session Initiation Protocol (Voice over IP); http://www.elektronik-kompendium.de/sites/net/0905111.htm, January 2005.

Kapitel 3

Vertrauen und Identität im Internet

Joerg Steinmann, Fabian Gubler, Iwan Stierli

Unsere Ausarbeitung des Themas Vertrauen und Identität im Internet möchte einen systematischen Überblick gewähren und die wichtigsten Aspekte in 4 Bereiche Unterteilen. Dazu gehören eine Übersicht über die verschiedenen webbasierten Anwendungen mit den dazugehörigen Interaktionspartner, eine Einteilung der verschiedenen Aspekte in soziale und technische und den Bereich der Rechtsprechung, staatliche Einflussnahme und Zertifikate. Der erste Teil, die Übersicht über die verschiedenen Anwendungen mit den zugehörigen Interaktionspartner dient dazu, aufzuzeigen wo und warum Vertrauen und Identität eine so wichtige Rolle spielen. Der zweite Teil mit den subjektiven, sozialen Aspekten hat das Ziel die verschiedenen Mechanismen und Wirkungsweisen in der menschlichen Wahrnehmung und Entscheidungsfindung im Medium Internet zu erklären. Der dritte Teil zeigt technische Möglichkeiten auf, um die Problematik der Sicherheit und somit des Vertrauens zu entschärfen. Der vierte Teil beschäftigt sich mit der Rechtslage im Internet, die Möglichkeit von Zertifikationsprozessen und Standardisierungen und die Rolle des Staates bzw. Wirtschaftsraums. Dabei ist die Gesetzgebung als Instrument zu verstehen, das eine Brücke zwischen sozialen, technischen und gesellschaftlichen Aspekten aufspannt und so Vertrauen schenken soll, wo dies aufgrund der Interaktion zwischen verschiedenen Partner notwendig ist.

Inhaltsverzeichnis

3.1 Einleitung

Diese Arbeit wurde im Rahmen des Seminars „Internet Economies“ erarbeitet und soll das Thema der Identität und des Vertrauens im Internet ausleuchten, die Problematik diskutieren und mögliche Lösungskonzepte aufzeigen.

Ein wesentliches Anliegen dieser Arbeit ist es die Vertrauensfrage nicht einfach global zu bearbeiten, sondern aufzuzeigen, wo denn Vertrauen eine wesentliche Rolle spielt und dass Identität und Vertrauen sehr eng zusammen spielen.

3.1.1 Wo spielen Vertrauen und Identität eine Rolle

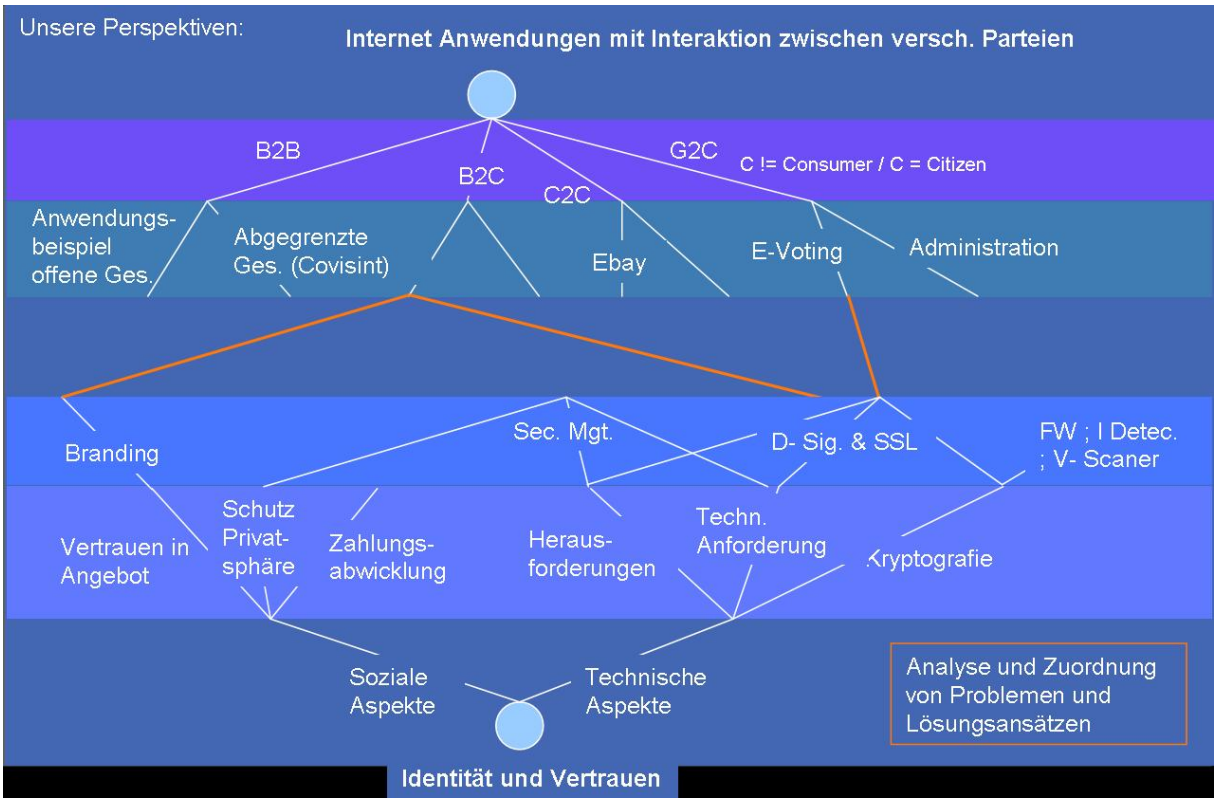

Abbildung 3.1: Übersicht- unsere Perspektiven

Wenn wir die verschiedenen Beziehungen zwischen zwei oder mehr Parteien anschauen, die im Internet zu Transaktionen oder zu einem Informationsaustausch führen, so stellen wir fest, dass es im wesentlichen vier Grundtypen gibt:

1. B2B: beinhaltet sämtliche Interaktionen zwischen zwei geschäftstreibenden Parteien, also zwei Firmen, bei der die eine oder beide Leistungen voneinander beziehen, kooperieren oder zusammen in Relation innerhalb einer Wertschöpfungskette stehen (Beispiele: Kooperationsplattformen für Fahrzeugentwicklung wie Covisint, Supply-Chain-Plattformen, Discrete Collaboration, distributed Enterprise etc.)

2. B2C: umfasst sämtliche Verhältnisse zwischen einer geschäftstreibenden Partei und einem Endkunden (in der Regel Privatperson) (Beispiele: Webshops, Administrationsplattformen, elektronische Auftragserfassungstools, Mass Customization, etc.)

3. C2C: beschreibt sämtliche Interaktionen zwischen Privatpersonen, die mit einander Transaktionen oder Informationsaustausch via Web abwickeln (Beispiele: Auktionen, Tauschplattformen, elektronische Marktplätze)

4. G2C (hier steht C nicht für Consumer sondern für Citizen): steht für Regierungs-und Bürger-Beziehungen, die übers Web abgewickelt werden (Beispiele: elektronischer Urnengang, Bürgeradministration via Webformulare, etc.)

Bei den drei Typen B2B, B2C und C2C gibt es weitere Unterteilungen. Man unterscheidet zwischen offenen und geschlossenen Gesellschaften. Zum Beispiel sind auf einer Supply-Chain-Plattform in der Regel nur spezifische Zulieferungspartner zusammengeschlossen und entsprechend zugelassen, die sich gegenseitig zur Zusammenarbeit ausgewählt haben und sich auf spezifische Standards und Normen geeinigt haben. Das wäre eine geschlossene Gesellschaft, die durch Verträge unter den Mitglieder schon viele Fragestellungen und Unsicherheiten bereinigt hat. Bei offenen Gesellschaften ist der partizipierende Kreis von Interaktionspartner nicht eingeschränkt und somit sind die Kenntnisse über den „Anderen“ in der Regel nicht existent oder nur durch kleine Transaktionserfahrungen vorhanden.

3.1.2 Offene Fragen, Bedenken und Hemmnisse

Aus den oben aufgeführten Web-Anwendungsmöglichkeiten ergeben sich schnell viele Fragestellungen, welche die persönlichen oder firmeneigenen Sicherheitsbedürfnisse betreffen. Dazu ein paar Beispiele:

Besonders zwischen Anbieter und Abnehmer von Dienstleistungen, Produkten und anderen Transaktionsgüter ist die Frage nach dem gültigen Gerichtsstand ein zentraler Punkt. Aber ebenso wird sich der Anbieter fragen: „Bekomme ich mein Geld?“ und vor allem in welchem Zeitraum, oder hält sich die andere Partei an die getroffenen Abmachungen? - Und der Abnehmer wird Bedenken haben, ob das erworbene Gut in seinem Land zu den üblichen Spezifikationen (Stromnetz, Funknetz, GSM-Standard, etc) kompatibel ist oder in der Anwendung unbedenklich ist und die erwartete Wirkung auch zeigen wird (Medikamente, Kosmetika, etc.).

Zwischen zwei kooperierenden Unternehmen wird sich mindestens der eine Partner sicher Gedanken machen, ob die Geschäftsgeheimnisse und strategischen Informationen geheim gehalten werden können cite9. Aber ebenso können unterschiedliche Security-Bedürfnisse oder -Standards zu hemmenden Bedenken führen.

Ein besorgter Bürger wird sich Gedanken machen, ob seine Stimme richtig gezählt wird und seine Annonymität gewahrt wird oder ob der Nachbar heraus findet, wie er abgestimmt hat. Im Bereich der Bürgeradministration via Webplattformen wird wohl die Annonymität und die Datengeheimhaltung ebenso ein wichtiger Anspruch sein vom Benutzer.

Um diese Fragestellungen und Bedenken zu überwinden braucht es in das Medium 'Internet' und die andere Partei oder Parteien ein unbedingtes Vertrauen. Daher ergibt sich die zentrale Frage, wie denn dieses Vertrauen konstruktiv aufgebaut und wie Missbrauch verhindert werden kann. In dieser Arbeit unterteilen wir deshalb der Systematik wegen dem Aspekt Vertrauen in soziale und technische.

3.2 Vertrauen

3.2.1 Definition Vertrauen

> Zuviel Vertrauen ist häufig eine Dummheit, zuviel Misstrauen ist immer ein Unglück. Johann Nestroy (1801-1862)

> Vertrauen bedeutet zu erwarten, dass gesetzte Erwartungen erfüllt werden, obwohl ein Risiko besteht, dass die Erwartungen nicht erfüllt werden. (Frässdorf) [3]

Vertrauen wird im alltäglichen Sprachgebrauch oft verwendet und man ist sich allgemein einig, dass Vertrauen ein sehr wichtiges Konstrukt ist. Auch in der Wissenschaft hat der Vertrauensbegriff in den letzten Jahren an Interesse und Bedeutung gewonnen. Es gibt jedoch keine allgemeingültige Definition des Begriffs Vertrauen. Vielmehr wird in der Literatur der Vertrauensbegriff unterschiedlich definiert und verschiedene Aspekte des Vertrauens werden betont.

Die etlichen Versuche Vertrauen zu definieren, können als Annäherungen an die unterschiedlichen Dimensionen des Begriffs betrachtet werden. Der Vertrauensbegriff wird somit immer mit einer Mehrdimensionalität verbunden.

Zusammengefasst kann gesagt werden, dass Vertrauen eine Voraussetzung für das soziale Zusammenleben der Menschen ist. Vertrauen reduziert die Komplexität der Welt und ermöglicht uns Verträge, Austauschbeziehungen, Kooperationen und Versprechen einzugehen und eröffnet uns eine Vielzahl von Handlungsmöglichkeiten. Vertrauen birgt auch immer ein gewisses Risiko in sich, weil man dadurch seine Verletzlichkeit und seine Schwächen seinem Gegenüber preisgibt, mit der Erwartung, dass diese Situation nicht ausgenutzt wird. Ob Vertrauenshandlungen wegen der Moral, reiner Sympathie, dem Eigeninteresse wie zum Beispiel dem Erhalt des guten Rufes, aus rationalen Kosten-Nutzen- Abwägungen oder aus Angst vor Sanktionen eingehalten werden, darüber sind sich die Soziologen und Philosophen nicht einig. Dass der Schaden bei einem Vertrauensbruch grösser ist, als der Vorteil aus dem eingegangenen Vertrauen, wird von den meisten unterstützt. Dieser Schaden kann durch die Suche nach Informationen über den Vertrauensnehmer reduziert werden, doch bei vollständiger Information kann man kaum mehr von Vertrauen reden. Das heisst, bei einer Vertrauenssituation geht man immer von einem Informationsmangel aus, was in der Praxis meistens der Fall ist. Vertrauen ist zukunftsorientiert und basiert auf einer gewissen Vertrautheit. Persönliche Erfahrungen können das Vertrauen beeinflussen, indem jemand, der bei einer Vertrauenshandlung zu Schaden gekommen ist,

wahrscheinlich in Zukunft nicht mehr so schnell anderen Vertrauen schenken kann. Ob der Vertrauensprozess unbewusst geschieht oder ob man dabei Zeit hat, die Vertrauensvergabe abzuwägen, wird in der Literatur nicht schlüssig beantwortet. Möglicherweise gibt es beide Arten der Vertrauensbildung, wobei es auf die jeweilige Situation ankommt. Der Vertrauensbegriff hängt sehr mit dem Menschenbild zusammen. Nimmt man an, dass Menschen von Natur aus nur aus einem rationalen Eigeninteresse handeln, dann sind Sanktionen nötig, um das Vertrauen aufrecht zu erhalten. Falls man hingegen an die Moral und das Gute im Menschen glaubt, ist Vertrauen einfacher zu bilden. Vermutlich liegt die Realität irgendwo in der Mitte dieser beiden Ansichten.

Vertrauen bedeutet Zuversicht. Oftmals stellt es sich durch positive Erfahrungen her oder wird durch Merkmale wie gesellschaftlichen Status, Sachverständigkeit, Unabhängigkeit von Parteiinteressen usw. konstituiert und unterstützt. Vertrauen kann als eine Relation in der Sachdimension definiert werden, also als Tatsache, dass ein Mensch einem Ding, einer Technik, einem Medium (ver)traut, es als authentisch ansieht. Oder aber man betrachtet Vertrauen als Kalkulation, die sich auf zukünftige Zusammenarbeit richtet:

> [...] trust is the calculation of the likelihood of future cooperation and is a defining feature of virtual cooperation. As trust declines, people are increasingly unwilling to take risks an demand greater protections against the probability of betrayal [...] (Ratnasingham).

Im E-Commerce ist genau diese Abschätzung des Risikos zentral. Damit Vertrauen bestehen kann, ist ein entsprechender Schutz vor der Wahrscheinlichkeit einer Gefahr unbedingt nötig.

3.2.2 Reputationssysteme

> Reputation ist gewissermassen die öffentliche Information über die bisherige Vertrauenswürdigkeit eines Akteurs. Sie spielt zum Beispiel bei Kooperationen eine grosse Rolle, wenn keine nachprüfbaren Informationen über die Leistungsfähigkeit potentieller Partner vorliegen oder im Internet-Handel, wenn durch die Informationen im Web keine Rückschlüsse auf die Qualität der Leistung möglich sind. (Einwiller 2003, S. 92)[11]

Reputation widerspiegelt die öffentliche Meinung und Wahrnehmung eines Anbieters. Es handelt sich um eine positive Bewertung eines Akteurs, welche unter den Mitgliedern eines sozialen Netzwerks verbreitet wird. Einwiller definiert Reputation in ihrer Arbeit als „der gute Ruf eines Reputationsobjektes, welcher aus der sozial vermittelten Einstellung Dritter gegenüber selbigem resultiert.“

Reputation hängt stark mit der Kommunikation zusammen, denn ohne diese könnte die Reputation gar nicht entstehen. Falls ein Konsument nicht auf seine eigenen Erfahrungen und Beobachtungen zurückgreifen kann, muss er sich auf die Informationen Dritter, das heisst auf die Reputation des Anbieters verlassen. Anhand dieser Informationen werden dem Anbieter gewisse Vertrauenseigenschaften zugeschrieben. Reputation kann ein

wichtiger Faktor für den Vertrauensaufbau eines Anbieters sein. Wenn jemand anders positive Erfahrungen mit einem Internetanbieter gemacht hat und dies weitererzählt oder in einem Forum davon schreibt, dann fällt es dem Konsumenten leichter dem Anbieter zu vertrauen. Reputation wird somit nicht nur durch die Kommunikation mit dem Anbieter selber, sondern auch durch den Austausch mit anderen Konsumenten aufgebaut. Wichtiges Instrument zum Reputationsaufbau ist die Kommunikation. Verschiedene Medien, wie Newsgroups, Internet-Chats, Verbraucherportale, Beschwerdewebseiten, Presse etc. unterstützen diesen Austausch. Sicherheits- und Datenschutz-Gütesiegel, die von anbieterunabhängigen, vertrauenswürdigen Instanzen nach festgelegten Kriterien vergeben werden, können auch zum Aufbau von Reputation beitragen. Solche Instanzen werden „Trusted Third Parties“ genannt und ermöglichen Vertrauensbeziehungen mit per se unbekannten Transaktionspartnern einzugehen. In diesem Zusammenhang kommt es zu einer Verschiebung des Vertrauensobjektes, denn es muss nicht mehr direkt dem Interaktionspartner sondern der Drittpartei vertraut werden. [1] Unterschieden werden muss zwischen den Anbietern, die das Internet als zusätzlichen Vertriebskanal benützen und in der realen Welt einen Einkaufsladen besitzen und solche, die nur im Internet präsent sind. Die erstgenannten können die schon erworbene realweltliche Reputation bei ihrem Onlineauftritt übernehmen und vom Transfereffekt profitieren. Sogenannte Newcomer müssen zu Beginn viel in vertrauensbildende Massnahmen investieren. Eine Möglichkeit, den Reputationsaufbau zu beschleunigen, besteht in einer Kooperation mit Partnern, die schon eine hohe Reputation besitzen.

Markennamen werden oft mit der Reputation in Verbindung gebracht. Marken können positive und starke Bedeutungsinhalte in den Köpfen vieler Konsumenten hervorrufen. Eine Marke kann als „ein in der Psyche des Konsumenten verankertes, unverwechselbares Vorstellungsbild von einem Produkt oder einer Dienstleistung“ (Einwiller 2003, S. 101) [11] angesehen werden. Eine Marke vermittelt einem Kunden ein Image über eine Unternehmung. Die Marke, das Image und somit auch die Reputation korrelieren positiv miteinander. Die Reputation kann somit zusätzlich anhand einer Marke aufgebaut werden.Abschliessend kann gesagt werden, dass der Reputationsaufbau eine wichtige Voraussetzung für einen Anbieter ist, damit er gegenüber den Konsumenten vertrauenswürdig wirkt. Mit Hilfe der Reputation wird beim Kunden die Unsicherheit reduziert und Vertrauen aufgebaut. Somit wurde gezeigt, dass die Reputation ein wichtiges Vertrauenskonstrukt ist.

Im Folgenden Abschnitt wird auf fünf Kategorisierungen eingegangen, die von Diekmann/Wyder [6] eingeführt wurden und anhand denen die Einbettung der Online-Reputation in die allgemeine Form der Reputationssysteme ersichtlich ist:

Informelle Reputation in sozialen Netzwerken
Hierbei geht es um die Reputation, die eine Person einer anderen Person in einer bestimmten Umgebung zuschreibt. Zum Beispiel kann dies sein, wenn ein Student durch andere Studenten für einen Posten in den Studentenrat empfohlen wird.

Reputation durch Markennamen
Ein Produkt ist in der Perzeption der Konsumenten mit einem Image, einer Reputation, eng verknüpft. Die Hersteller kreieren nun vor Nachahmung geschützte Produkte, die

[1] Weiteres dazu kann im Kapitel Auslagerung der Vertrauensfrage nachgelesen werden.

überall in gleicher Qualität angeboten werden. Als Beispiele können hier Coca Cola oder McDonalds genannt werden.

Experten-Rating
Hierbei wird Reputation durch eine Fachperson einem Produkt zugeschrieben. Diese Form der Reputationszuschreibung wird zum Beispiel bei Produktetests durch Verbraucherorganisationen durchgeführt.

Konsumentenrating
Beim Konsumentenrating wird einem Produkt durch die Benutzer, wobei diese auch Laien darstellen können, eine gewisse Reputation verliehen. Es gibt hierbei auch im Internet institutionalisierte Formen, bei denen Käufern nahe gelegt wird, eine Rezension über eine Ware zu verfassen. Teilweise werden diese Testberichte durch Anreize belohnt, teilweise werden sie auch ohne finanzielle Anreize vergeben, wie zum Beispiel bei Büchern von Amazon.

Reputationsverfahren von Internet-Auktionen
Die Reputationsverfahren von Online- Auktionen, die institutionalisiert und hochsystematisch ablaufen, können als eigener Zweig der Reputationssysteme betrachtet werden.

Als eine weitere Differenzierung der Reputationssysteme kann zwischen negativen und positiven Reputationssystemen unterschieden werden: Von positiven Reputationssystemen spricht man, wenn ein zufriedener Kunde seine guten Erfahrungen mit dem Handelspartner öffentlich kund tut. Im Gegensatz dazu spricht man von negativer Reputation, wenn unzufriedener Kunde seinem Ärger über das missglückte, aus seiner Sicht unbefriedigend abgewickelte Geschäft, in seinem Umfeld negative Werbung macht. Generell geht man davon aus, dass Individuen in negative Reputation wesentlich mehr Energie stecken, als in positive. Oder anders ausgedrückt: ein unzufriedener Kunde macht seinem Unmut Luft, in dem er andere potentielle Kunden von einer Geschäftsabwicklung mit „seinem“ Geschäftspartner abrät.

3.2.3 Entmaterialisierung des Geldes

Die Entmaterialisierung von Geld ist ein Vorgang, der bereits um 1900 von Georg Simmel, einem deutschen Soziologen, thematisiert wurde. Es ist von einem Prozess die Rede, der zu dem Zeitpunkt begann, als „Handel“ in seiner ursprünglichen Form des Tauschs von Naturalwerten zum ersten Mal stattfand. Der Blick in die Vergangenheit lässt erkennen, dass Medialisierung nicht nur ein Begriff unserer schnelllebigen, konsumorientierten Moderne ist, sondern dass bereits während der Zeit des Jahrhundertwechsels Medienwandel zu Verlagerung von Vertrauen führte. Es stellt sich folgende Frage: Inwiefern hat sich das Vertrauen in (elektronischen) Handel durch die Entmaterialisierung des Geldes verändert?

Um dies zu beantworten, soll der Vertrauenstransfer in der Entwicklung des Mediums Geld erläutert und aufgezeigt werden. Der Beginn dieser Entwicklungskette bildet die achetypische Form des Handelns, das Tauschen. Danach wird die Medialisierung als Faktor eingeführt, der speziell in der entsinnlichten Welt der Online-Kommunikation Vertrauensgewinnung deutlich erschwert. Im Folgenden werden Aspekte diskutiert, die nicht nur die

Entwicklung von Geld aufzeigen, sondern vor allem die Ansprüche an die Vertrauensgewinnung eines neuen Mediums ersichtlich machen. Das (digitale) Geld hat den höchsten Grad der Entmaterialisierung erreicht. Worin bestehen nun die Bausteine des Vertrauens und der Glaubwürdigkeit in der heute bereits grossflächig etablierten Handelsform des E-Commerce? Was ist nötig, um eine Generalisierung dessen zu ermöglichen?

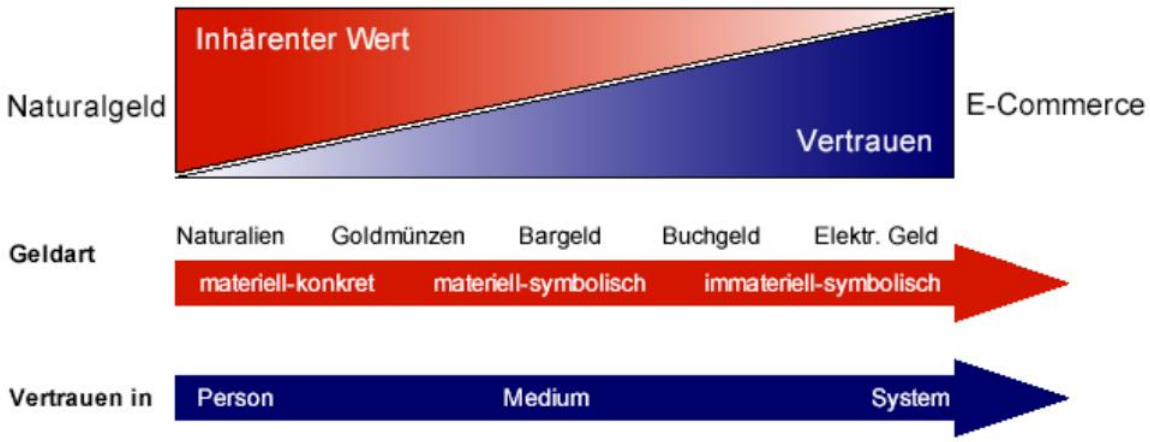

Abbildung 3.2: Entmaterialisierung des Geldes [4]

Abbildung 3.2 versucht darzustellen, in wie weit die Entmaterialisierung des Geldes Einfluss auf die Vertrauensfrage nimmt. Im linken Teil der Graphik finden wir uns zeitlich in einer Epoche, als Geld, wie wir es heute kennen, noch unbekannt war. Tauschhandel ist eher der Terminus, welcher auf die damaligen Geschäftstätigkeiten passt. Wollte ein Mensch beispielsweise ein lebendiges Schwein kaufen, so tat er dies durch einen Tauschhandel mit beispielsweise Tierfellen. Der Wert der Felle ist in diesem Fall materiell-konkret gegeben. Vertrauen brauchte man bei solchen Arten von Geschäften höchstens in seinen Handelspartner, ausserdem war eine direkte Prüfung der Geldart möglich. Dieser Tauschhandel verlor mit der Zeit allerdings an Bedeutung, da er mit sehr viel Aufwand verbunden war. Langsam entwickelte sich das Medium Geld, erst in Form von Goldmünzen (welche an sich selbst noch einen Wert besassen), später dann in Form von Münzen und Noten, welche bloss symbolisch für einen gewissen Wert standen. Damit veränderte sich die Vertrauensfrage. Ab sofort war nicht mehr der Handelspartner der primäre Dreh- und Angelpunkt der Vertrauensfrage, sondern das Medium (Goldmünzen), später dann das gesamte System, welches die Bar- beziehungsweise Buchgeld zur Verfügung stellte. In der aktuellsten Betrachtungsweise, ist der inhärente Wert des Geldes gänzlich verschwunden, der Wert des elektronischen Geldes definiert sich nur noch durch das Vertrauen, welches man dem System entgegenbringt. Diese doch nicht unbedeutende Wandlung, was Geld anbelangt und die damit verbundenen Konsequenzen beim Vertrauen im Internet, werden in den nächsten zwei Kapitel noch ein wenig ausführlicher Betrachtet.

3.2.4 Vertrauen in der Zahlungsabwicklung

Wohl oder übel spielt in der heutigen Zeit Vertrauen erst eine übergeordnete Rolle, wenn allfällige Konsequenzen aus Vertrauensverlust oder -missbrauch monetärer Natur sind. Aus diesem Grund soll in diesem Kapitel auf das Vertrauen im Internet gekoppelt an Zahlungen näher betrachtet werden. Gängigste Mittel zur Bezahlung von Dienstleistungen oder Waren im Internet sind zur Zeit:

Vorauskasse
Bei dieser Art der Bezahlung trägt der Käufer das grösste Risiko. Wird von ihm beispielsweise verlangt, bei einer Bestellung den Warenwert im Voraus auf ein Bankkonto des Verkäufers zu überweisen, fehlt ihm bei einem allfälligen Nicht-Erhalt der Ware die Möglichkeit, gegen den Verkäufer zu intervenieren. Anzufügen gilt, dass diese Aussage sich auf Summen bezieht, bei welchen sich der Beizug von Rechtshilfen nicht lohnt.

Nachnahme
Unter Nachnahme versteht man die Bezahlung direkt beim Erhalt der Ware. Konkret bezahlt man dem Postboten, sobald dieser die bestellte Ware liefert, den vereinbarten Preis. Die Post wiederum liefert das Geld dann dem Verkäufer weiter. Problematisch an dieser Methode sind die hohen Gebühren, welche die Post für diese Dienstleistung erhebt und ausserdem die Problematik, dass nicht reelle Güter (wie beispielsweise ein Software-Download), welche nicht durch die Post geliefert werden, mit dieser Methode nicht bezahlt werden können.

Rechnung
Einigen sich die Parteien darauf, die Bezahlung der Waren mittels einer Rechnung zu erledigen, sieht die Problematik in etwa, bis auf umgekehrte Vorzeichen, wie bei der Bezahlung mittels Vorauskasse aus. Sollte nämlich der Käufer die Rechnung nicht bezahlen, so bleibt dem Verkäufer nichts anderes übrig, als den Käufer mehrmals zu mahnen und anschliessend ein Inkasso zu eröffnen. Dies lohnt sich aber bloss in wenigen Fällen, da die streitbare Summe eine Inkasso-Eröffnung meist nicht lohnt.

Kreditkarten
Da sich die Kreditkarte im Gegensatz zu allen oben genannten Möglichkeiten der Bezahlung durch eine einfache Handhabung, wenig zusätzliche Kosten und eine relativ grosse Sicherheit für beide Parteien auszeichnet, ist sie zur Zeit wohl das gängigste und beliebteste Mittel, im Internet zu bezahlen. Um ein besseres Verständnis des nächsten Kapitels zu erlangen, ist in den folgenden paar Sätzen kurz ausgeführt, wie die Bezahlung einer Kreditkarte genau funktioniert.

Kreditkarten wurden Anfangs dieses Jahrhunderts eingeführt, um den Leuten ein einfaches Mittel zum bargeldlosen Zahlen zur Verfügung zu stellen. Heutzutage sind Kreditkarten weltweit akzeptiert, besitzen Millionen von Akzeptanzstellen und Millionen von Kreditkarteninhabern. Die bekanntesten Marken sind zur Zeit MasterCard, VISA und American Express, welche in verschiedensten Ausführungen daherkommen und von ihren Trägern teilweise als Statussymbol getragen werden (Gold- oder Platinkarten). Die Zahlungsabwicklung ist aber, unabhängig vom Kartenprodukt, stets die selbe.

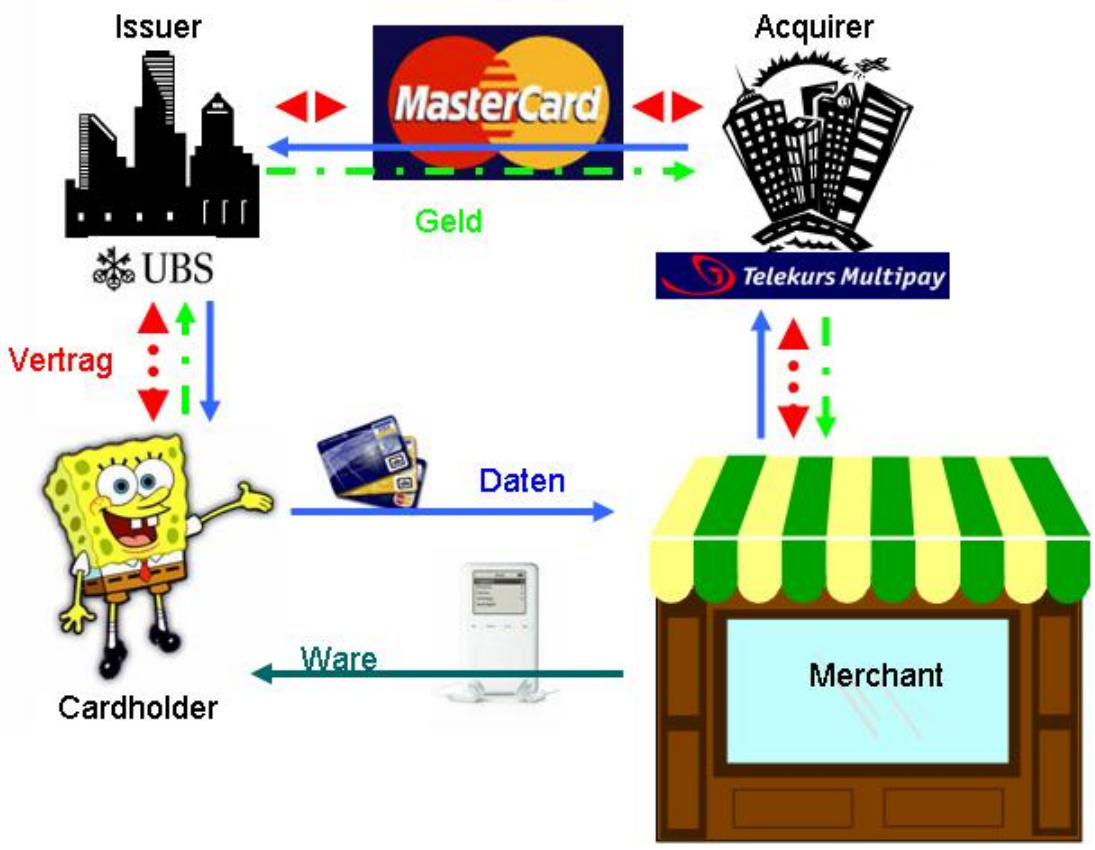

Abbildung 3.3: Kreditkartenzahlungssystem

Will ein Karteninhaber (Cardholder) eine Ware oder ein Gut mittels Kreditkarte bezahlen, so muss er sich erst mal eine solche Besorgen. Dazu schliesst er mit dem sogenannten „Issuer“ einen Vertrag ab und kommt so zu seiner Karte. Damit er nun aber in einem Geschäft (oder in unserem Fall: im Internet) wirklich mit der Karte bezahlen kann, muss dies Geschäft seinerseits einen Vertrag mit einem sogenannten „Acquirer“ abschliessen, welcher es ihm ermöglicht, Kreditkarten zu akzeptieren. Issuer und Acquirer unterstehen dabei einem sogenannten „Card Scheme“, im Beispiel in Abbildung 3.3 ist dies MasterCard International [16, 17]. Das Card Scheme gibt die Regeln vor, wie Issuer und Acquirer miteinander zu geschäften haben.

Soll es nun zu einem Kauf kommen, präsentiert der Cardholder im Geschäft seine Kartendaten und erhält dafür die Waren. Das Geschäft wiederum liefert die Kreditkartendaten an den Acquirer, dieser an den Issuer. Dann fliesst das Geld vom Issuer über den Acquirer zum Geschäft. Der Karteninhaber selbst hat zu diesem Zeitpunkt noch kein Geld ausgeben müssen, er geniesst beim Issuer Kredit, daher der Name „Kredit-Karte“. Am Ende eines sogenannten Rechnungslaufes werden dem Karteninhaber vom Issuer die Informationen über seine Einkäufe präsentiert und der Totalbetrag eingezogen. Damit schliesst sich der Kreis und der Einkauf ist abgeschlossen.

3.2.5 Auslagerung der Vertrauensfrage

In diesem Kapitel nun schliesst sich der Kreis, welcher in den vorhergehenden zwei Kapiteln eröffnet wurde, und der Bezug zum Thema dieser Arbeit wird erstellt. Und zwar ging die Überlegung in Richtung, dass bei Transfer von Geld im Internet das Vertrauen eine übergeordnete Rolle spielt, und es sich deshalb besonders lohnt, Zahlungsweisen im Internet kurz zu durchleuchten. Dabei wurde vor allem auf die gängigste Zahlungsart eingegangen, die Zahlung mittels Kreditkarte. In diesem Kapitel werden Vermutungen dargelegt, weshalb genau die Kreditkarte trotz gewissen Sicherheitsbedenken der Karteninhaber immer noch das beliebteste Zahlungsmittel ist.

Grund dafür ist die gegenseitige Auslagerung der Vertrauensfrage. Während bei allen anderen Zahlungsmitteln eine direkte Eins-zu-Eins-Beziehung zwischen den Partnern entsteht, wird die Verantwortung bei Kreditkartenzahlungen abgeschoben. Der Karteninhaber beispielsweise muss die Vertrauenswürdigkeit des Verkäufers nicht mehr prüfen, er verlässt sich bei seiner Zahlung darauf, dass bei allfälligen Fehlern der Karten-Issuer dies Problem schon regeln wird. Zahlt er also beispielsweise eine Ware im Internet und wird diese nicht geliefert, so teilt er dies dem Issuer mit und hofft, dass dieser in seinem Sinne mit den Händler Kontakt aufnehmen wird.

Auf der anderen Seite ist der Verkäufer davon überzeugt, dass wenn er eine Zahlung mittels Kreditkarte bekommt, welche er über seinen Acquirer authorisieren lässt, dass er auf jeden Fall zu seinem Geld kommt, da der Issuer ja für die Bonität der Karteninhaber verantwortlich ist. Bei einer allfälligen Insolvenz eines Karteninhabers also liegt das Verlustrisiko beim Issuer der Karte.

Zusammenfassend kann also davon gesprochen werden, dass die Teilnehmer einer geschäftlichen Tätigkeit im Internet darauf bedacht sind, die eigentlich Frage nach dem Vertrauen in den Handelspartner Drittorganisationen zu überlassen. Über Sinn und Unsinn einer solchen Verlagerung (die eigentliche Problematik ist ja dabei noch nicht gelöst) lässt sich unedlich weiter diskutieren.

3.2.6 Die Problematik des gläsernen Menschen

Das Recht auf den Schutz der Privatsphäre gilt mittlerweile in allen internationalen Abkommen über Menschenrechte als ein fundamentales Recht, das auch in den meisten Verfassungen der Länder explizit oder implizit anerkannt wird, selbst wenn natürlich die Wirklichkeit dem nicht entspricht. Der Schutz der Privatsphäre ist sogar zu einem der wichtigsten Menschenrechtsprobleme des modernen Zeitalters geworden. Schwierig ist allerdings eine genaue Definition dessen, was als Recht auf Privatsphäre betrachtet wird. Es beschränkt sich aber nicht nur auf Datenschutz, sondern damit zieht man allgemein eine Grenze, wie weit die Gesellschaft sich in die Angelegenheit der Einzelnen einmischen darf.

Besonders die neuen Technologien können die Rechte des Einzelnen einfacher untergraben, weil sie die Erhebung, Speicherung, Verarbeitung und Verknüpfung von vielen Daten für Sicherheitsbehörden, aber auch für privatwirtschaftliche Interessen erheblich leichter

machen. Neben Überwachungskameras, Sammlung von medizinischen und genetischen Daten, Smart Cards oder biometrischen Mitteln zur Identifizierung ist vor allem das Internet Gegenstand wachsender Überwachung, während die Überwachungsmöglichkeiten regelmässig, selbst in den demokratischsten Ländern, missbraucht werden. In vielen Ländern fällt die Gesetzgebung hinter den technologischen Fortschritt zurück, wodurch Lücken entstehen. Mehr als 90 Länder überwachen illegal die Kommunikation von politischen Gegnern, Menschenrechtsaktivisten, Journalisten und Gewerkschaftern. Und selbst in Ländern mit strengen Datenschutzgesetzen wie in Schweden oder Norwegen besitzt die Polizei umfangreiche Dateien über Bürger, die weder angeklagt noch eines Verbrechens beschuldigt wurden. Weitere Gründe, die die Aushöhlung der Privatsphäre begünstigen sind beispielsweise die Globalisierung, welche die geographischen Grenzen des Datenstroms zerstört. Die Konvergenz der Techniken überwindet die Grenzen zwischen Systemen.

Verschiedene Gesetze und Richtlinien geben inzwischen Weisungen vor, wie der Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten und beim freien Datenverkehr zu handhaben ist. Im Mittelpunkt steht dabei das Recht, über den Entstehungsort von Daten in Kenntnis gesetzte zu werden, falsche Daten korrigieren zu können, Einspruch gegen eine illegale Verarbeitung von Daten einlegen zu können und Daten nicht mitteilen zu müssen. Am wichtigsten ist, dass der Betroffene explizit der kommerziellen und staatlichen Verwendung von Daten zustimmen muss.

Offen ist allerdings noch der Schutz der Privatsphäre am Arbeitsplatz in vielen Ländern. Angestellte in nahezu allen Ländern könnten durch die Vorgesetzten überwacht werden, wobei gesetzliche Bestimmungen meist ziemlich schwach sind, da die Möglichkeit der Überwachung oft als Bedingung der Einstellung gilt.

Abschliessend gilt an dieser Stelle zu bemerken, dass die Problematik des so genannten „gläsernen Menschen“ trotz staatlichen Versuchen, das Problem in der Gesetzgebung zu lösen, in den nächsten paar Jahren an Bedeutung gewinnen wird. Grund dafür mag sein, dass in der heutigen Dienstleistungs-Zeit Informationen als äusserst wertvolles Gut gelten und dem ihrem Besitzer oft direkte Vorteile monetärer Natur bringen. Und so lange irgendwelche, wenn auch halblegale Möglichkeiten bestehen, an Informationen über Menschen zu gelangen, werden diese auch ausgeschöpft werden. Der beste Schutz gegen diese Persönlichkeits-Spionage bietet noch immer die Igel-Taktik, in dem man sich so weit wie möglich abkapselt und nirgends im Netz Informationen über sich selbst preis gibt [13].

3.2.7 Identität

> Unter Identität im klassischen Sinne versteht man das Bewusstsein, sich von anderen Menschen zu unterscheiden (Individualität) sowie über die Zeit (Kontinuität) und verschiedene Situationen (Konsistenz) hinweg - auch für die Umwelt erkennbar - dieselbe Person zu bleiben. (Nicola Döring)[12]

Das Konzept der Identität ist eng verbunden mit Kommunikation, Persönlichkeitsrecht und Sicherheit. Die Übereinkunft der Kommunikation via Email beispielsweise wäre unmöglich ohne gewisse Konventionen für die Identifikation untereinander. Aber wie überall

besteht natürlich auch hier die Möglichkeit des Missbrauchs welche sogar ernste Konsequenzen nach sich ziehen können.

Jedoch ist gerade auch das Verbergen/Verändern der Identität eine der größten Attraktionen des Internet, welche es Leuten ermöglicht, in Rollen zu schlüpfen, welche sie sich im echten Leben niemals einzunehmen trauten. Diesbezüglich sei hier explizit erwähnt, dass die Bewusste Verschleierung der eigenen Identität im Internet nicht zwangsläufig und auch eher selten einen „kriminellen Hintergrund“ hat. Häufiger der Fall ist, das aus dem Vorgaukeln einer fremden Identität eine Art „Spiel“ entsteht, bei dem alle Parteien wissen, oder zumindest ahnen, dass ihr virtuelles Vis-à-Vis eine neue Identität angenommen hat. Gut zu beobachten sind solche Handlungsweisen in so genannten Chat-Räumen. Doch trotz diesen (geduldeten) Identitätsverschleierungen beruhen heutzutage viele Geschäftmodelle und Verhandlungsbasen auf verlässlichen und sicheren Techniken der Identifikation. In den folgenden, technik-basierten Kapiteln wird auf solche Techniken und Möglichkeiten näher eingegangen.

Normalerweise ist die Identität im Internet aus einer Reihe von Gründen ein amorphes und fast nichtexistentes Konzept. Ein Grund ist der ständige Fluß im Cyberspace , in dem Leute regelmäßig auftauchen und verschwinden wobei eine zeitweilige Abwesenheit von der ”Gemeinde” nicht wirklich bemerkt wird. Die meisten Leute erinnern sich an Gesichter und Stimmen, die Hauptmerkmale der Identifikation im ”echten” Leben. Die verzweigten und kryptischen Folgen von Buchstaben und Ziffern auf die die meisten Email Adressen reduziert sind sind nicht mal annähernd zu behalten und weit entfernt von der eindeutigen Identifikation eines Lebewesens, das multiple Accounts auf multiplen Rechnern überall auf der Welt haben kann [12].

3.3 Technische Aspekte

Der nachfolgende Abschnitt soll die Rolle der Technik im Zusammenhang mit dem Vertrauensaufbau etwas beleuchten und konkretisieren. Zudem werden aktuelle Techniken eingeführt, welche für eine vertrauliche und „sichere“ Kommunikation übers Internet heutzutage unumgänglich sind.

3.3.1 Bedeutung der Technik für Vertrauen

Damit ein System als vertrauenswürdig angesehen wird, muss es eine Reihe von Erwartungen und Anforderungen erfüllen. Dabei stellt die Sicherheit die grössten Bedenken bezüglich des elektronischen Handels dar [2]. Die Gründe für ein solches Misstrauen liegen auf der Hand. Der technologische Fortschritt geht immer rascher voran. Die Systeme werden immer komplexer und somit auch unüberschaubarer. Die älteren Generationen halten mit diesem Fortschritt selten stand. Man spricht hier von einem „digital divide“ (Der Unterschied zwischen den Wissenden und Technikversierten gegenüber den Unwissenden, Konservativen, Technikabgeneigten.) Diese Spaltung wird mit zunehmender Digitalisierung immer grösser.

Neben einer Zunahme des Entwicklungstempos nehmen auch die täglichen Negativmeldungen und Warnungen von Bedrohungen im Internet zu (Viren, Trojaner, Phishing,...). Diese Meldungen sind nicht gerade vertrauensweckend, sondern eher abschreckend. Ein weiterer Misstrauen bildender Aspekt ist, dass Angriffe als solche oftmals gar nicht sofort erkannt werden.

Grundsätzlich muss betont werden, dass technische Sicherheit eine notwendige aber keine hinreichende Bedingung für Vertrauen im elektronischen Handel ist. Die Sicherheit welche von einem System gegeben wird hängt vom Individuum ab. Sicherheit in elektronischen Systemen lässt sich nicht messen und ist nicht sichtbar. Die Sicherheit ist darum immer relativ und situationsbezogen. Trotz all diesen Schwierigkeiten gibt es verschiedene Mittel und Techniken um die empfunden Sicherheit zu erhöhen. 100%ige Sicherheit wird es leider nie geben. Wie aber die Sicherheit lässt sich durch bestimmte Mittel und Techniken erhöhen?

Wie diese Techniken konkret aussehen und was für Anforderungen diese Systeme erfüllen müssen für den elektronischen Handel, ist Gegenstand der nachfolgenden Abschnitte.

3.3.2 Technische Anforderungen

Das Vertrauen in die Technik spielt eine Schlüsselrolle für den elektronischen Handel. Wie bereits erwähnt muss ein System diverse Anforderungen und Bedingungen erfüllen, damit ein System als sicher erachtet wird und entsprechend Vertrauen aufbaut. Die wichtigsten Anforderungen an ein System werden in den folgenden fünf Punkten beschrieben.

1. **Verfügbarkeit**
   Unter Verfügbarkeit versteht man die sichere und dauernde Benutzbarkeit eines Systems. Um dies zu erreichen braucht es hauptsächlich ausreichende und adäquate Ressourcen. Probleme in diesem Bereich lassen häufig auf ungenügende Hardware schliessen. Die Gewährleistung der Verfügbarkeit ist aber nicht nur von der Hardware abhängig sondern auch von vielen technischen Massnahmen. Weil dies zum Teil Gegenstand von anderen Gruppen ist, soll an dieser Stelle nicht weiter auf das Problem der Verfügbarkeit eingegangen werden. (vgl. Distributed Denial of Service Attack)

2. **Vertraulichkeit**
   TCP/IP-Pakete werden normalerweise unverschlüsselt übermittelt. Somit sind sowohl übermittelte E-Mail-Mitteilungen als auch WWW-Seiten grundsätzlich von jedem Router aus, bei dem die Pakete vorbeikommen, lesbar. Da es auch nicht möglich ist, den Weg der Pakete bei der Übermittlung im Voraus zu bestimmen, ist die Vertraulichkeit von über das Internet übermittelten Daten ohne zusätzliche Hilfsmittel nicht gewährleistet.

3. **Authentizität**
   Authentizität liegt dann vor, wenn ein Dokument auf die als Aussteller angegebene Person zurückgeführt werden kann. Die Absenderadressen von E-Mail-Nachrichten können in aller Regel durch den Anwender in der E-Mail-Software frei eingestellt

werden und sind damit beliebig fälschbar. Hinzu kommt, dass auch das Benutzerkonto, von dem aus eine E-Mail- Nachricht verschickt wurde, nicht immer eruierbar ist. Die Authentizität von E-Mail- Nachrichten lässt sich daher oftmals nicht nachweisen. Auch im World Wide Web ist die Authentizität von Daten nicht sicher eruierbar: So ist es durch Manipulation eines Routers möglich, Abfragen auf einen falschen WWW-Server umzuleiten.

4. **Datenintegrität**
   Der Begriff der Datenintegrität ist ziemlich schwer zu definieren. Gemeinhin wird gesagt, Daten seien dann integer, wenn sie nicht durch Unberechtigte verändert wurden. Sämtliche auf einem Computer gespeicherten Daten können schon durch einen wenig versierten Anwender problemlos verändert werden. Dies gilt auch für früher empfangene oder versandte (d.h. im Ausgangsordner gespeicherte) E-Mail-Nachrichten oder für eigentlich nicht zur Nachbearbeitung vorgesehene Dokumentenformate, wie etwa das Portable Document Format (PDF). Leicht veränderbar sind grundsätzlich ebenso die Randdaten, also etwa das Erstellungs- oder Veränderungsdatum einer Datei oder Nachricht.
   Auf nur einmal beschreibbare Medien wie CD-R geschriebene Daten können zwar als solche nicht mehr geändert werden, hier besteht jedoch auf die nahe liegende Möglichkeit, einen Datenträger mit leicht verändertem Inhalt einfach neu zu erstellen. Im Weiteren können Dateien auf einem derartigen Datenträger oftmals unsichtbar gemacht und durch neue ersetzt werden, ohne dass dies beim Aufruf der Inhaltsanzeige für den Datenträger erkennbar wäre.
   Bei der Datenübermittlung über das Internet kann ein Angreifer aufgrund dessen Routing- Konzepts die über einen für ihn zugänglichen Router geleiteten Datenpakete nicht nur lesen, sondern auch verändern. Weder bei der Speicherung von Daten auf normalen oder einmal beschreibbaren Datenträgern noch bei der Übermittlung von Daten über das Internet ist die Datenintegrität damit gewährleistet. Schliesslich ist noch darauf hinzuweisen, dass die Unterscheidung zwischen Authentizität und Integrität streng genommen nicht sinnvoll ist. Denn wenn ein Text als nicht authentisch gilt, hat die Prüfung der Integrität keinen Sinn, und wenn ein Text nicht integer ist, ist er gleichzeitig auch nicht mehr authentisch.

5. **Nichtabstreitbarkeit**
   Nichtabstreitbarkeit (Non-Repudiation) liegt dann vor, wenn der Aussteller eines Dokumentes sich dessen Inhalt aufgrund des Nachweises von dessen Authentizität und Integrität anrechnen lassen muss. Beim materiellen Briefverkehr würde ein eingeschriebener Brief diese Anforderung erfüllen. Beim elektronischen Briefverkehr ist dies eines der grössten Probleme. Sprüche wie: „Was für ein E-mail? Ich habe keines erhalten!“ sollten allen bekannt vorkommen [1].

3.3.3 Digitale Signatur

Die digitale oder elektronische Signatur ist ein digitales Siegel, das die Identität des Ausstellers und die Unveränderheit digitaler Dokumente nachweist und es dem Aussteller verunmöglicht, seine Urheberschaft an diesen Dokumenten zu bestreiten. Sie erfüllt damit

bei der elektronischen Kommunikation ähnliche Funktionen wie die Handunterschrift auf Papier. Bereits 1994 wurde im schweizerischen Nationalrat die rechtliche Gleichstellung von digitaler Signatur und Handunterschrift gefordert. Damit sollte ermöglicht werden, schriftformbedürftige Rechtsgeschäfte auch auf elektronischem Wege abzuschliessen. Im Frühling 2000 erliess der Bundesrat die Zertifizierungsdiensteverordnung, welche versuchsweise die zum Einsatz elektronischer Signaturen nötige Sicherungsinfrastruktur regeln sollte. Auf Anfang 2003 soll mit dem Bundesgesetz über Zertifizierungsdienste im Bereich der elektronischen Signatur (ZertES) ein Erlass in Kraft treten, der diese Verordnung ersetzt und nebst der Infrastruktur weitere Bereiche abdeckt, darunter die ursprünglich geforderte Gleichstellung der digitalen Signatur mit der Handunterschrift sowie die Benutzung von digitalen Signaturen soll nun Schritt für Schritt durch die Einführung von verschiedenen Techniken und Hilfsmittel erläutert werden.

Kryptographie

> Definition: Kryptographie bezeichnet die Anwendung mathematischer Algorithmen zur Ver- und Entschlüsselung elektronischer Daten beim Senden und Empfangen[...] (Zasterbox.de)[18]

Die geschilderten Schwierigkeiten hinsichtlich Vertraulichkeit, Authentizität, Datenintegrität sowie Nichtabstreitbarkeit können durch Kryptographie angegangen werden. Kryptographie ist eine Disziplin der Informatik, die sich mit der Entwicklung und Bewertung von Verfahren zur Verschlüsselung geheimer Daten befasst. Das Handwerk der Kryptographie fand schon früh Anwendung in unserer Gesellschaft. Damals wie auch heute noch, kommt die Kryptographie bei den Militärs sowie auch bei den Liebesbrief- Schreibern zum Einsatz. Mittlerweile hat sich die Kryptographie zu einer eigenen Wissenschaft entwickelt und stark an Bedeutung zugenommen.

Symmetrische Verschlüsselung

Das herkömmliche Verständnis von Verschlüsselung basiert darauf, dass zur Verschlüsselung und Entschlüsselung jeweils der gleiche geheime Schlüssel eingesetzt wird. Man spricht hier auch von symmetrischen Verfahren.

Die am weitesten verbreitete symmetrische Verschlüsselungsmethode ist heutzutage DES (Data Encryption Standard). Seine Sicherheit schwindet jedoch zusehends aufgrund der rasch fortschreitenden Entwicklung in der Computertechnik, weshalb seit Oktober 1999 zunächst der Einsatz einer dreifachen Variante (Triple-DES) empfohlen und hernach Rijndael[2] als neuer Standard definiert wurde.[3] Symmetrische Verschlüsselung setzt voraus, dass der geheime Schlüssel vor Aufnahme der Kommunikation auf sicherem Wege, also etwa via Telefon oder persönlich, ausgetauscht werden kann. Genau diese Voraussetzung ist

[2]Sprich: *reindohl.*

[3]Zum Ganzen FOX, DES, 736; Stefan LUCKS/Rüdiger WEIS, Der DES-Nachfolger Rijndael, DuD 12/2000, 711ff.; vgl. auch Patrick BRAUCH (pab.), Belgisches Krypto-Verfahren setzt sich durch, c't 22/2000, 84.

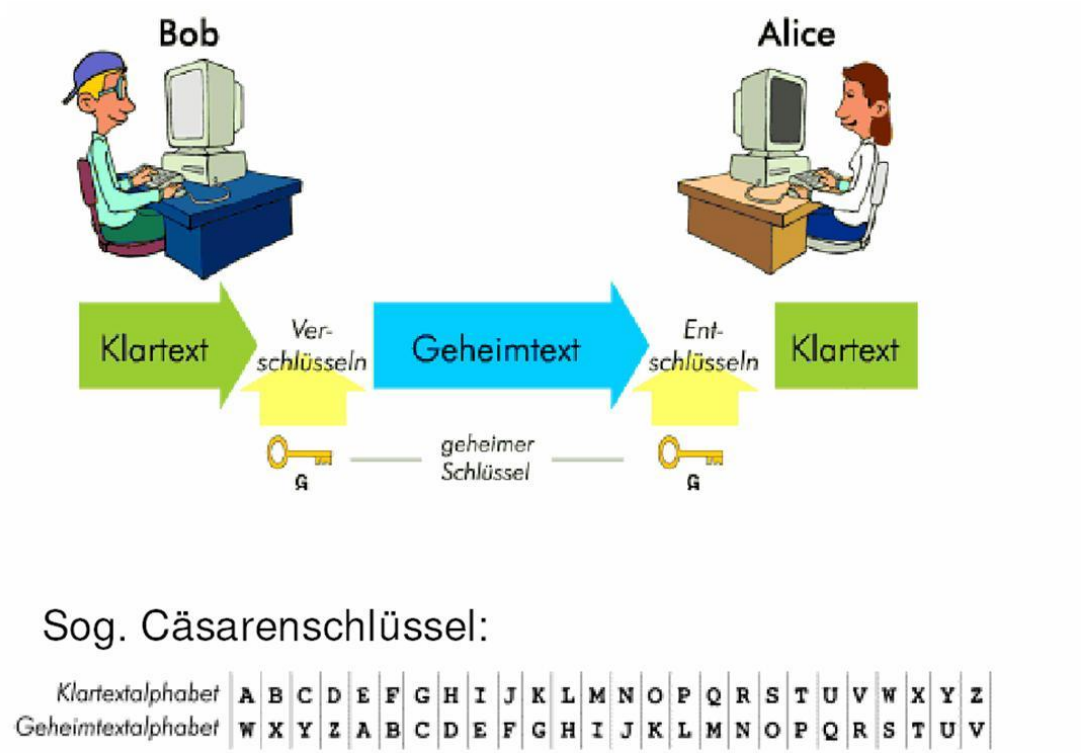

Abbildung 3.4: symmetrische Verschlüsselung

aber bei der Kommunikation über das Internet nicht gegeben: Schon beim ersten Kontakt soll möglichst eine vertrauliche Verbindung zwischen den beiden Kommunikationspartnern aufgebaut werden können. Daher müssen alternative Verfahren eingesetzt werden.

In Abbildung 3.4 wird die symmetrische Verschlüsselung illustriert. Das einfachste Beispiel für die symmetrische Verschlüsselung ist die so genannte Cäsarenverschlüsselung. Dabei verschiebt man jeden Buchstaben des Alphabet um x Stellen. Dabei stellt x den Schlüssel dar.

Grundsätzlich ist es so, dass in einem Verschlüsselungssystem der Verschlüsselungsalgorithmus bekannt ist und nur der Schlüssel ist unbekannt. Alleine das Unbekanntsein des Schlüssels soll verhindern, dass das System nicht geknackt werden kann.

Asymmetrische Verschlüsselung

Das derzeit am weitesten verbreitete Verfahren, welches für den Datenaustausch über offene Netzwerke wie das Internet eingesetzt werden kann, wurde bereits 1978 von den US-Amerikanern Ron RIVEST, Adi SHAMIR und Leonard ADLEMAN entwickelt. Das nach seinen Erfindern benannte RSA-Verfahren basiert - gleich wie digitale Signaturen - auf einem Schlüsselpaar pro Kommunikationspartner, wovon ein Schlüssel zur Ver- und der andere zur Entschlüsselung eingesetzt wird. Der Entschlüsselungsschlüssel bleibt geheim, der Verschlüsselungsschlüssel wird veröffentlicht. Weil die Verschlüsselungsschlüssel öffentlich sind, ist kein geheimer Schlüsselaustausch mehr nötig. Die Verwendung eines Paares

unterschiedlicher Schlüssel führte zum Begriff des asymmetrischen Verfahrens; asymmetrische Verschlüsselung bildet nebst der digitalen Signatur den zweiten Anwendungsbereich der Public-Key-Kryptographie. Wie soll etwas mit einem Schlüssel verschlüsselt werden

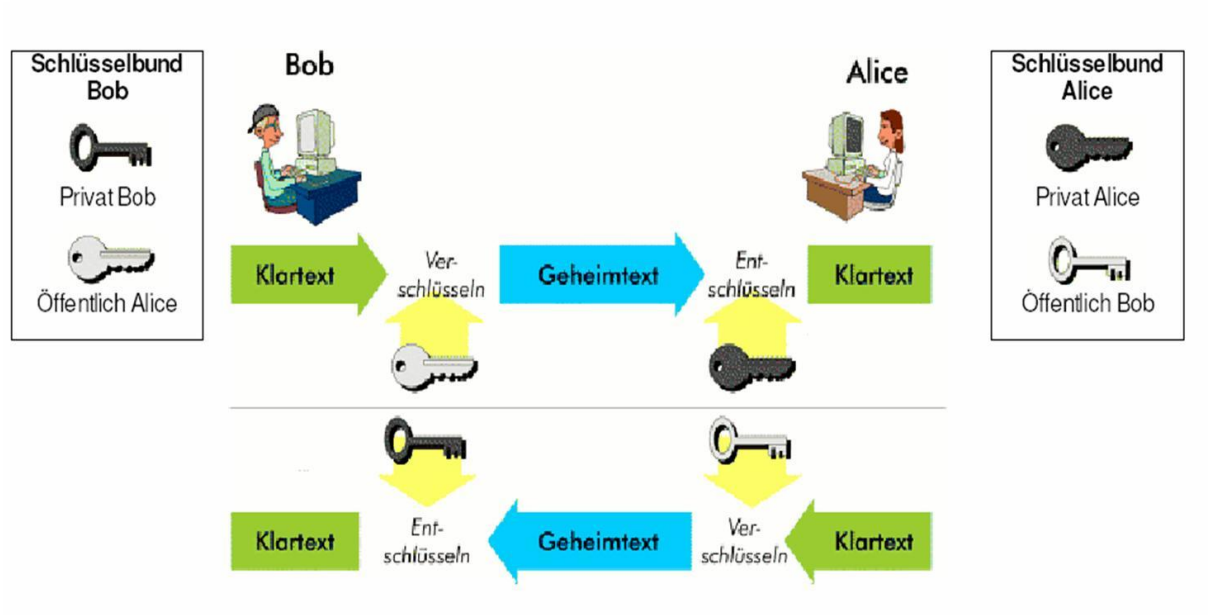

Abbildung 3.5: asymmetrisch Verschlüsselung [15]

und danach ist man nicht mehr in der Lage dies zu entschlüsseln? Für einen logisch denkenden Mensch ist dies kaum sinnvoll! Nun um dies zu verstehen braucht man einiges an mathematischem Verständnis. Kryptographische Schlüssel sind miteinander mathematisch korrelierende, mehrhundertstellige Zahlen, die - vereinfacht gesagt - aufgrund ihrer Grösse nicht erraten werden können. Mit einem dieser Schlüssel werden digitale Signaturen gesetzt (Signierschlüssel) und mit dem anderen überprüft (Prüfschlüssel). Der Signierschlüssel ist geheim zu halten, weil jedermann, der auf ihn Zugriff hat, damit Signaturen setzen kann. Er wird daher auch geheimer Schlüssel oder private key genannt. Der Prüfschlüssel hingegen muss allen potentiellen Adressaten zur Prüfung von Signaturen zur Verfügung stehen, weshalb er zu veröffentlichen ist. Er wird deswegen auch öffentlicher Schlüssel oder public key genannt. Verfahren, die auf solchen Schlüsselpaaren basieren, werden gemeinhin als Public-Key-Kryptographie oder asymmetrische Kryptographie bezeichnet. Obwohl die Schlüssel mathematisch korrelieren, darf es nicht möglich sein, den Signierschlüssel aus dem Prüfschlüssel oder einer Signatur zu errechnen. Bei der Erzeugung des Signierschlüssels ist dafür zu sorgen, dass niemand ausser dem späteren Halter zu diesem Zugang erhält. Dies wird im einfachsten Fall dadurch erreicht, dass der Halter das Schlüsselpaar auf seinem eigenen Computersystem erzeugt, wo der Signierschlüssel sofort durch ein Passwort geschützt wird.

Hashfunktion

Um eine Digitale Signatur zu erstellen, erwartet der Signieralgorithmus als zu signierenden Input eine Zeichenkette bestimmter Länge. Bevor eine digitale Signatur gesetzt werden

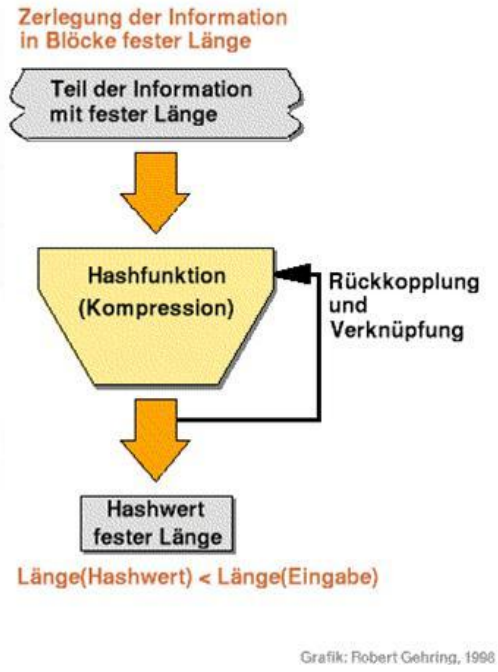

Abbildung 3.6: Hashfunktion [15]

kann, muss aus der zu signierenden Nachricht, die ja beliebig lang sein kann, als erstes eine verkürzte Form erzeugt werden. Diese muss für jeden Text wieder anders aussehen, denn es darf nicht passieren, dass verschiedene Texte dieselbe Signatur ergeben. Diese verkürzte Form der Nachricht wird Hashwert genannt. Vgl hierzu Abbildung 3.6
Ein Hashwert ist demnach eine Zahl mit einer genau bestimmten Anzahl Stellen, beispielsweise 128 Bit, die für jede Nachricht einen unterschiedlichen Wert aufweist. Man spricht aus diesem Grund auch von einem digitalen Fingerabdruck eines Dokuments; der Hashwert kann das Dokument vertreten. Eine weitere Anforderung an eine solche Hashfunktion ist, dass sie nicht umkehrbar sein darf. Aus einem Hashwert darf die Originalnachricht nicht wiederhergestellt werden können [15].

Signieren

Anhand den bis hierher erläuterten Grundwerkzeugen zur Digitalen Signatur, sollte der Signierprozess anhand der Abbildung 3.7 klar sein! Für weiteren Erläuterungen vgl. [14]

BEMERKUNG: Vielfach konzentriert man sich zu stark auf die asymmetrische Verschlüsselung. Nun im Gegensatz zur Geheimhaltungsverschlüsselung verwendet man für die Digitale Signatur den privaten Schlüssel des Senders und nicht den öffentlichen Schlüssel des Empfängers. Soll das signierte Dokument geheim (verschlüsselt) übermittelt werden, so muss das Dokument nach dem Signieren zusammen mit der Signatur mit dem öffentlichen Schlüssel des Empfängers noch verschlüsselt werden.

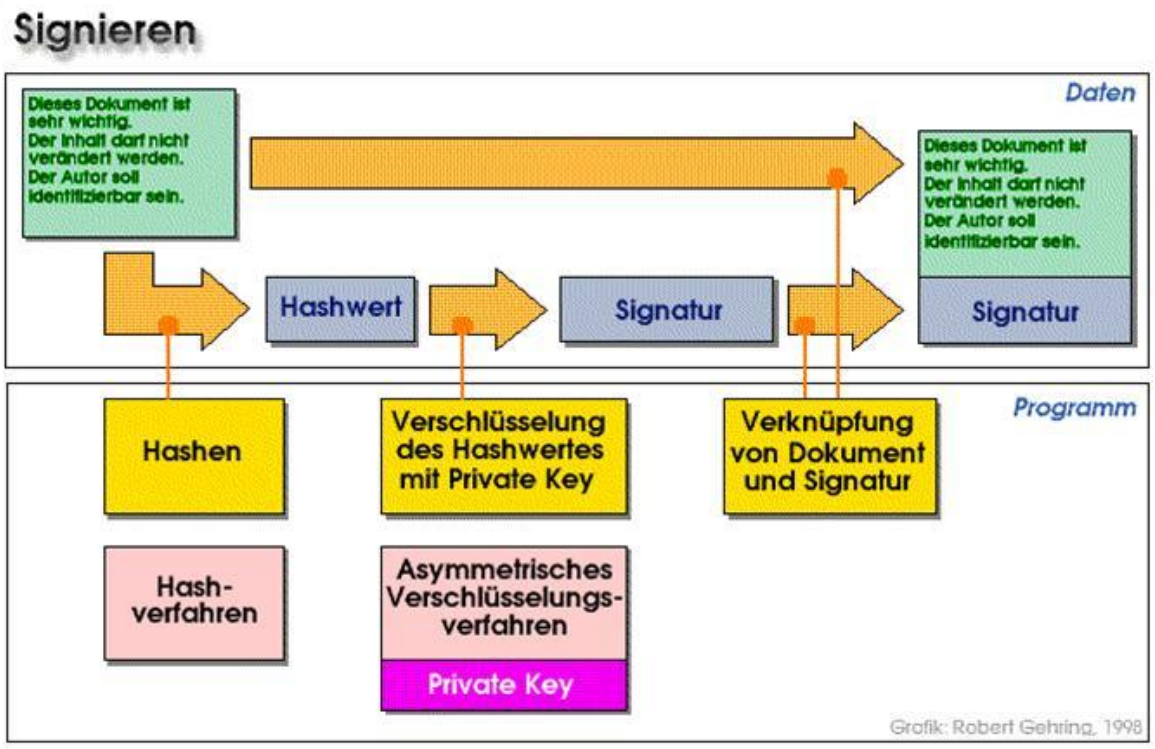

Abbildung 3.7: Prozessdarstellung für die Signaturerstellung [15]

Prüfen der Signatur

In dieser Sektion soll zuerst der Ablauf der Überprüfung der digitalen Signatur erklärt werden und auf mögliche Schwachpunkte und Sicherheitsbedenken hinweisen.

1. **Allgemein**
   Die Prüfsoftware prüft, ob Nachrichteninhalt, Prüfschlüssel und Signatur zusammenpassen. Wurde der Text seit dem Setzen der Signatur verändert, oder passen Signier- und Prüfschlüssel nicht, so schlägt diese Überprüfung fehl. Die Software zeigt dem Anwender das Prüfresultat durch ein Symbol an. Marktübliche E-Mail-Software, die digitale Signaturen unterstützt, führt diese Prüfung normalerweise beim Empfang signierter Nachrichten automatisch durch. Die Gesamtheit der zum Überprüfen einer Signatur eingesetzten Soft- und Hardware wird Signaturprüfeinheit oder Prüfeinheit genannt. Es handelt sich hier üblicherweise um einen herkömmlichen Computer.

2. **Vergleich der Hashwerte**
   Genau gesehen erstellt die Software des Empfängers zum Prüfen der Signatur einerseits den Hashwert der übermittelten Nachricht und ermittelt andererseits mittels des Prüfschlüssels den zuvor durch den Absender mit dessen privaten Schlüssel verschleierten Hashwert. Stimmen die beiden Hashwerte überein, so ist die Signatur gültig.

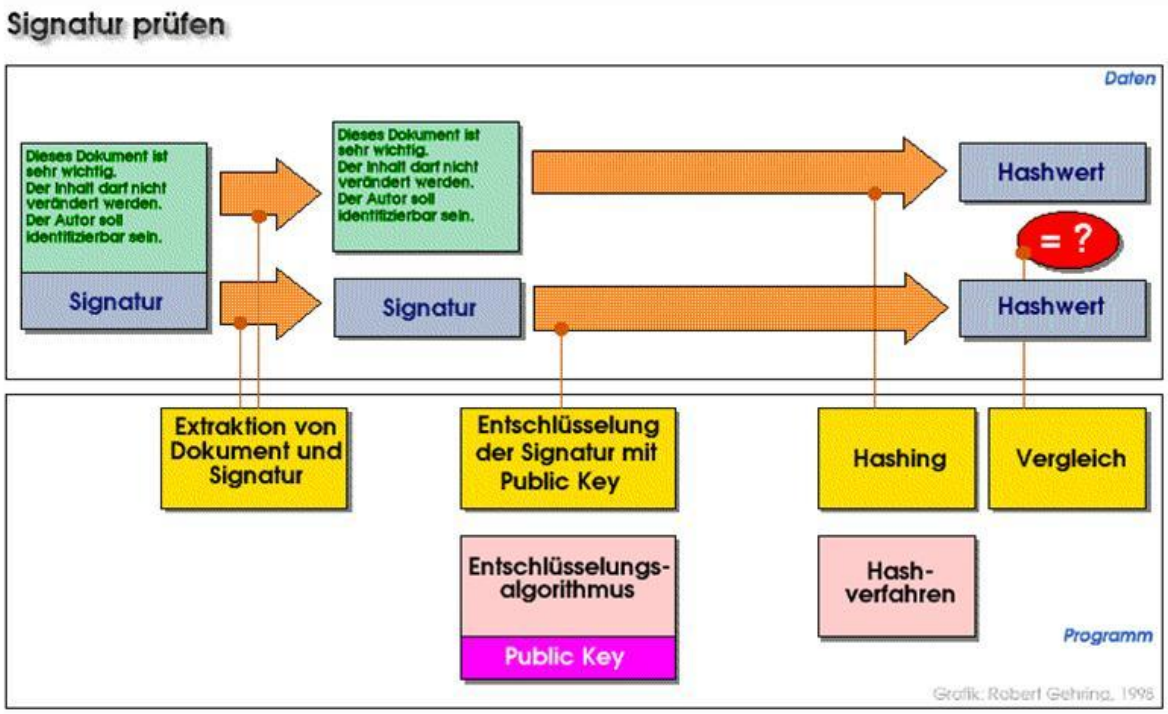

Abbildung 3.8: Prüfen der Signatur [15]

3. **Überprüfung der Zertifikate**[4]
   Im Weiteren ist die Gültigkeit der involvierten Zertifikate zu überprüfen, das Wurzelzertifikat sogar anhand eines nicht-digitalen Referenzwertes (z.B. seines publizierten Hashwertes). Eine grosse Schwierigkeit liegt hierbei im fehlenden Know-how der Anwender oder in deren Bequemlichkeit. Die Signaturprüfeinheit sollte die Zertifikatsdaten oder zumindest Teile davon daher nicht nur auf Initiative des Anwenders hin und erst nach einigen Mausklicks, sondern bereits im Hauptfenster anzeigen. Die Abfrage der Gültigkeit von Zertifikaten über die Certificate Revocation List CRL kann online und softwaregesteuert sowie meist auch „von Hand“ durch einen Besuch auf der Website des Zertifizierungsdiensteanbieters erfolgen. Um den Anwender zu entlasten, sollte die Signaturprüfeinheit die CRL bei jeder Signaturprüfung automatisch und ohne Benutzerinteraktion konsultieren; dies ist bei gängiger Software jedoch noch selten der Fall.
   Sehr komplexe Probleme ergeben sich ferner bei einer Kompromittierung des Signierschlüssels des Zertifizierungsdiensteanbieters und einer Sperrung des Wurzelzertifikats; daher wird heute auf eine sehr hohe Sicherheit der entsprechenden Signierhardware Wert gelegt.

4. **Angriff durch Täuschung des Empfängers**
   Der Empfänger kann darüber getäuscht werden, welche Bereiche eines Dokumentes signiert sind. Problematisch ist dieser Aspekt besonders bei strukturierten, mit mehreren Signaturen versehenen Dokumenten. Bei der Auswahl der Software ist daher darauf zu achten, dass jeweils das gesamte Dokument unterzeichnet wird bzw. bei einer nur teilweisen Unterzeichnung der Empfänger auf diesen Umstand hingewiesen

[4] Zertifikate werden im nächsten Unterkapitel genauer erläutert.

wird.
Schliesslich ist es auch möglich, das Prüfsystem zu manipulieren. Weil bei der Prüfung keine geheimen Daten (wie Signierschlüssel) zur Anwendung gelangen, läuft die Prüfung regelmässig auf einem herkömmlichen Computersystem, das gegen Angriffe nur schlecht abgeschirmt werden kann [1].

3.3.4 Zertifikate/SSL

Der Einsatz digitaler Signaturen sowie der Aufbau von verschlüsselten Internetverbindung setzt voraus, dass die im Umlauf befindlichen Schlüssel aus Sicht die vertrauenden Dritten sicher ihren Haltern zugeordnet werden können. Werden die Schlüssel via Internet übermittelt, so kann der Empfänger einer Nachricht aber zunächst nicht sicher sein, wem der verwendete Prüfschlüssel wirklich gehört. Wie dieses Problem versucht wird zu lösen, ist Gegenstand dieses Kapitels. Zur sicheren Zuordnung von Prüfschlüsseln zu Personen gibt es drei Verfahren: Erstens können die Schlüssel dem Empfänger signierter Nachrichten natürlich vorab physisch übergeben werden - man spricht von Direct Trust -, zweitens kann mit einem Web of Trust und drittens mit Schlüsselzertifizierung (Hierarchical Trust) gearbeitet werden.

Web of Trust

Das Web of Trust ist ein Netz von Vertrauensbeziehungen, in dem Private, die ihrerseits über einen bereits beglaubigten Prüfschlüssel verfügen (so genannte „Introducer“) die Prüfschlüssel von Bekannten beglaubigen, d.h. durch ihre eigene digitale Signatur bezeugen, dass ein Schlüssel zu einer bestimmten Person gehört. Die „Introducer“ ermöglichen damit anderen Bekannten, diesen Prüfschlüsseln zu vertrauen, obwohl sie deren Halter nicht kennen. Dies funktioniert in einem überschaubar grossen Anwenderkreis gut. Die Sicherheit wird jedoch beeinträchtigt, wenn der Kreis der Schlüsselhalter grösser wird und eine eventuelle Überprüfung der Zuordnung eines Schlüssels nicht mehr in Betracht kommt. Das am weitesten verbreitete Verschlüsselungs- und Signierprogramm auf Public-Key-Basis, Pretty Good Privacy (PGP), basiert auf einem Web of Trust [1].

Hierarchical Trust

Bei der dritten und im Folgenden ausschliesslich betrachteten Variante wird das Problem der Zuordnung gelöst, indem die Zuordnung von Prüfschlüssel und rechtmässigem Halter des zugehörigen Signierschlüssels durch eine für die Kommunikationspartner vertrauenswürdige dritte Partei beglaubigt wird. Man spricht von Zertifizierung. Die dritte Partei wird gemeinhin Zertifizierungsdiensteanbieter (ZDA), Certification Service Provider (CSP) oder Certification Authority (CA) genannt. Die Beglaubigung erfolgt in einem ebenfalls elektronisch übermittelten Dokument, das Zertifikat genannt wird. Auch dieses wird durch eine digitale Signatur, diejenige des Zertifizierungsdiensteanbieters, gegen Fälschungen gesichert. Wohl aus der englischen Bezeichnung digital ID abgeleitet, wird das

Zertifikat teilweise auch als digitale Identität bezeichnet. Meines Erachtens müsste hier jedoch genau genommen von einem digitalen Identitätsausweis oder von einer digitalen ID gesprochen werden. Das Zertifikat enthält einige Daten zur Identifikation des Schlüsselhalters, den Prüfschlüssel, eine Seriennummer, den Namen des Zertifizierungsdiensteanbieters sowie meist einen Hinweis auf das so genannte Certification Practice Statement (CPS). In Letzterem dokumentiert der Zertifizierungsdiensteanbieter sein Vorgehen bei der Erstellung der Zertifikate (insbesondere wie die Identität des Schlüsselhalters überprüft wird). Eine sichere Identitätsprüfung kann etwa durch physisches Erscheinen bei einer Registrierstelle und Vorlegen eines amtlichen Dokumentes erfolgen. Teilweise wird aber auch bloss überprüft, ob die angegebene E-Mail-Adresse tatsächlich dem Antragsteller gehört (durch Zusenden eines Passwortes auf diese Adresse, welches der Antragsteller dann wieder einzugeben hat). Die Anschaffung eines Zertifikats mit Hardware kostet derzeit rund 150 Franken, sofern der Signierschlüssel auf einem Token gespeichert ist; für jedes Folgejahr sind erneut rund 50 Franken zu bezahlen [1].

Ablauf von Zertifikaten

Zertifikate werden regelmässig mit einem Verfalldatum versehen, was erlauben soll, der technischen Entwicklung etwa durch Anheben der Schlüssellängen zu folgen. Das Verfalldatum wird bei der Anzeige des Zertifikats durch die Signaturprüfeinheit regelmässig mit angezeigt.

Sperrung von Zertifikaten

Im Falle der Kompromittierung des Signierschlüssels, d.h., wenn dieser in die Hände Dritter gelangt ist, oder etwa wenn ein Mitarbeiter ein Unternehmen verlässt und eine entsprechende Vertretungsberechtigung verliert, ist das entsprechende Zertifikat bereits vor seinem Ablauf zu sperren. Teilweise wird auch von Widerruf oder Revokation gesprochen.

Kann ein Zertifikat gesperrt werden, so muss ein auf dieses vertrauender Dritter umgekehrt die Möglichkeit haben, ein Zertifikat zu einem bestimmten Zeitpunkt auf seine Gültigkeit hin zu überprüfen (dieses zu validieren). Üblicherweise veröffentlichen die Zertifizierungsdiensteanbieter hierzu auf dem Internet in regelmässigen Zeitabständen eine Zertifikatsrücknahmeliste (Certification Revocation List, CRL), d.h. eine Liste der gesperrten Zertifikate mit Informationen über Sperrzeitpunkt und -grund. Diese wird ihrerseits digital signiert, um die Vertrauenswürdigkeit zu gewährleisten und herkömmlicherweise als Ganzes heruntergeladen, um auch Offline-Zertifikatsprüfungen zu ermöglichen. Erfahrungsgemäss müssen jedoch jährlich rund 10 Prozent der ausgegebenen Zertifikate eines Anbieters vorzeitig gesperrt werden, was bei grösseren Anwenderzahlen sofort zu riesigen Datenmengen führt. Das CRL-System ist damit nicht beliebig skalierbar [1].

3.3.5 Probleme und Herausforderungen

In den vorherigen Kapiteln wurden zum Teil schon gewisse Problem und Herausforderungen vereinzelt aufgegriffen. Dieses Kapitel soll sich nun ausführlicher mit den möglichen

Angriffen und Problemen befassen. Ziel dieses Kapitels ist es, keine Aversion gegenüber der Technik zu schaffen, sondern vielmehr den Benutzer zu sensibilisieren. Viele Sicherheitsrisiken werden durch Unwissen von Benutzern hervorgerufen.

Kryptographische Sicherheit

Public-Key-Kryptographie basiert auf Problemen der Mathematik, für die es keine einfache und damit schnelle Lösungsmöglichkeit gibt. Solche Probleme sind etwa die Primfaktorzerlegung (eingesetzt im am weitesten verbreiteten RSA-Signaturverfahren), das Problem des diskreten Logarithmus, oder das „Travelling-Salesman- Problem“. Durch Wahl geeigneter Schlüssellängen (d.h. durch Vergrösserung des so genannten Schlüsselraumes, also der Anzahl aller möglichen Schlüssel) kann verhindert werden, dass ein Angreifer den Signierschlüssel innert für ihn nützlicher Frist durch Ausprobieren aller möglichen Schlüsselvarianten herausfinden kann (so genannte brute force attack). Ein Verfahren kann dann als sicher bezeichnet werden, wenn der Aufwand für eine Berechnung grösser wird als der zu erwartende Ertrag eines erfolgreichen Angriffs. Diese Sicherheit wird jedoch durch die kontinuierlich steigende Leistungsfähigkeit von Computern laufend vermindert, sodass die Verfahren und/oder die Schlüssellängen periodisch anzupassen sind. Die Verfahren unterscheiden sich teilweise erheblich bezüglich der für ausreichende Sicherheit notwendigen Schlüssellängen.

Zwar wurde bisher bei keinem dieser mathematischen Probleme bewiesen, dass keine einfache Lösungsmöglichkeit besteht, die eine Ermittlung von Signierschlüsseln in kurzer Zeit ermöglichen könnte. In Anbetracht dessen, dass die verbreitetsten Verfahren mittlerweile aber seit über 20 Jahren bekannt sind und jeglichen Angriffen standgehalten haben, werden diese Verfahren gemeinhin als sicher erachtet. Allenfalls wird die Entwicklung so genannter Quantencomputer der kryptographischen Sicherheit gegenwärtiger Public-Key-Verfahren dereinst ein Ende bereiten.

Alternativ zum Angriff auf die Public-Key-Verschlüsselung könnte ein Angreifer auch nach Kollisionen der verwendeten Hashfunktion suchen, um eine Nachricht zu erzeugen, die trotz unterschiedlichem Inhalt den gleichen Hashwert wie die Originalnachricht und damit die gleiche Signatur aufweist. Dies kann allerdings durch die Verwendung kryptographisch sicherer Hashfunktionen verhindert werden.

Die kryptographischen Algorithmen sind damit der stärkste Teil eines Systems zur Erstellung digitaler Signaturen. Ein potentieller Angreifer wird seine Zeit daher nicht mit Kryptoanalyse verschwenden, sondern sich viel eher auf andere Schwachstellen des Systems konzentrieren, also etwa versuchen, die Signierschlüssel durch Hacking zu beschaffen [1].

Präsentationsproblem

Eine grosse Schwierigkeit beim Design von Signiereinheiten liegt darin, dass die Signiersoftware zwar das vom Anwender gewollte Dokument auf dem Bildschirm anzeigen, im Hintergrund aber ein anderes Dokument signieren könnte. Um die Nichtabstreitbarkeit zu

gewährleisten, muss sichergestellt sein, dass angezeigter und signierter Text übereinstimmen („What you see is what you sign“).

Nebst dem Unterschieben falscher Dokumente etwa durch „bösartige Software“ (Viren, Trojanische Pferde u.dgl.) können etwa Textteile durch besondere Farbgebung unsichtbar gemacht (weisser Text auf weissem Grund), Zeichen auf unterschiedlichen Computersystemen unterschiedlich dargestellt (abweichende Zeichensätze),unsignierte Daten, die das signierte Dokument unbemerkt referenziert und anzeigt, verändert oder fein aufgelöste Grafikdetails aufgrund schlechter Auflösung der Anzeige verschluckt werden. Ohne die signierte Datei zu verändern führt all dies dazu, dass ein Dokument auf dem Prüfsystem anders angezeigt wird als auf dem Signiersystem. Man spricht hier gemeinhin vom Präsentationsproblem [1].

Vertrauenswürdige Hardware als Lösung
Das Präsentationsproblem ist bei fremden Signierumgebungen besonders gravierend. BIZER/HERDA schlagen daher vor, Signiervorgänge ausschliesslich in eigener gesicherter Systemumgebung durchzuführen, also etwa auf einem Personal Digital Assistant (PDA). Dies bewirkt zwar eine Steigerung der Sicherheit, Virenangriffe werden jedoch nicht ausgeschlossen. Um sicherzustellen, dass der Signaturhalter das richtige Dokument unterzeichnet, muss dieses durch vertrauenswürdige Hardware angezeigt und „gehasht“ werden. Derartige Hardwaresigniereinheiten, die eine eigene Anzeige aufweisen, werden Klasse-3-Geräte genannt. Der deutsche Hersteller Towitoko präsentierte mit dem „Chipdrive Monitor Kit“ eine Lösung, welche in das Monitorkabel eingeschleift werden kann und damit die Daten auch auf dem normalen Monitor sicher anzeigt. Andere Lösungen integrieren eine LCD-Anzeige mit wenigen Zeilen in den Kartenleser; sie kosten derzeit ungefähr zwischen 100 und 200 Franken. Leider eignen sich beide Lösungen nur für kleine Datenmengen; ungelöst bleibt im Weiteren das Problem, dass eine derartige Anzeige nur eine beschränkte Anzahl von Dokumentenformaten unterstützen kann.

Einige der geschilderten Schwierigkeiten können umgangen werden, indem das Format der signierten Daten vereinfacht wird. PORDESCH schlägt etwa eine Beschränkung auf ein „Primitivformat“ von 20 Zeichen pro Zeile und 20 Zeilen vor, das nur reinen Text ohne Formatierungen und Farben enthalten soll. Eine weitere Möglichkeit wäre, das Format der signierten Datei sowie bestimmte Systemparameter (etwa die verwendeten Schriftarten u.dgl.) im signierten Dokument festzuhalten, sodass zur Überprüfung nachweislich wieder ein gleiches System eingesetzt werden kann; dies kann jedoch unmöglich für die mannigfaltigen Konfigurationsmöglichkeiten eines herkömmlichen PC geschehen.

Bis zur Marktreife von wirklich sicheren Signiereinheiten für grössere Datenmengen und bis zur Etablierung von Standards zur Darstellung der signierten Dokumente dürfte noch einige Zeit verstreichen. Für grössere Datenmengen müssen daher vorderhand noch Softwarelösungen eingesetzt werden. Das Präsentationsproblem ist damit das schwerwiegendste Problem, das derzeit der technischen Absicherung der Nichtabstreitbarkeit durch digitale Signaturen noch entgegensteht.

Aufbewahrung und Schutz des Signierschlüssels

Die Handunterschrift ist durch körperliche und psychische Merkmale des Schreibenden geprägt; sie kann nicht von diesem getrennt werden und stellt quasi einen „unverlierbaren Schlüssel“ dar. Anders die digitale Signatur: Weil der Signierschlüssel bei Weitem zu lang ist, um in Erinnerung zu bleiben, muss er auf einem Datenträger gespeichert werden. Dieser Datenträger ist nicht in natürlicher Weise an dessen rechtmässigen Halter gebunden. Wenn ihn ein Angreifer unberechtigterweise entwendet, so ist er in der Lage, digitale Signaturen zu setzen, die mit der Originalunterschrift identisch sind.

Die Sicherstellung der alleinigen Kontrolle des berechtigten Halters über den Signierschlüssel ist folglich ein sehr wichtiges Postulat im Rahmen einer rechtlichen Regelung der digitalen Signatur. Die eingesetzten Produkte müssen dafür Gewähr bieten, dass der Signierschlüssel weder direkt durch Menschen noch indirekt durch bösartige Software (Viren u.dgl.) ausgespäht werden kann [1].

Soft- und Hardwareschutz
Oftmals wird der Signierschlüssel auf einem herkömmlichen Datenträger wie einer Diskette oder Festplatte gespeichert. Weil diesfalls sämtliche kryptographischen Operationen in Software auf dem Rechner ablaufen, soll dies in der Folge als Softwaresigniereinheit bezeichnet werden.

Die Speicherung des Signierschlüssels auf der Festplatte birgt das Risiko, dass dieser durch bösartige Software (Viren u.dgl.) ausgespäht oder sonstwie entwendet wird, ohne dass der Halter dies bemerkt. Zur Lösung dieses Problems kann der Signierschlüssel in einem separaten angriffsresistenten Chip gespeichert werden. Denkbar ist, den Chip fest in einem Computer zu verbauen oder ihn ähnlich einer SIM-Karte in ein Mobiltelefon zu stecken. Um den Chip nicht an ein bestimmtes Computersystem zu binden, wird er vielfach auch in ein so genanntes Token verbaut - etwa in eine Chipkarte oder in einen kleinen Stecker für den USB-Bus (USBToken). Derartige Lösungen werden in der Folge als Hardwaresigniereinheiten bezeichnet. Die auf dem Markt verfügbaren Tokens weisen sehr unterschiedliche Sicherheitsniveaus auf. Teilweise dienen sie bloss als Speicher für den Schlüssel, der wie ein normaler Datenträger gelesen werden kann und daher kaum mehr Sicherheit bietet als die Speicherung auf einem normalen Datenträger. Idealerweise können derartige Chips die Erzeugung der Schlüssel sowie die kryptographischen Operationen jedoch selbst durchführen. Diesfalls kann verhindert werden, dass der Signierschlüssel jemals in die unsichere Systemumgebung gelangt [1].

PIN oder Passwort
Verstärkter Schutz gegen den Diebstahl des Tokens oder das Ausspähen einer Schlüsseldatei wird in der Regel durch eine PIN85 oder ein Passwort erreicht. Um über den Schlüssel verfügen zu können, ist diesfalls nebst dem Zugriff auf die Schlüsseldatei bzw. (bei Hardwaresigniereinheiten) Besitz am Token auch Wissen nötig.

Ein Passwortschutz ist indessen oft nur beschränkt wirksam. Besteht das Passwort etwa aus wenigen Zeichen oder verwendet der Anwender ein umgangssprachliches Wort, so kann es mittels einer brute force attack, d.h. einer Angriffsmethode, bei alle möglichen Passwortvarianten durchprobiert werden, oder mittels einer Wörterbuchattacke, bei

der sämtliche Wörter eines elektronischen Wörterbuchs durchprobiert werden, ermittelt werden.

Im Gegensatz zu Softwaresigniereinheiten können Hardwaresigniereinheiten aber immerhin erlauben, eine bestimmte maximale Anzahl misslungener Versuche festzulegen, nach denen die Signierfunktion gesperrt wird.

Wird das Passwort über die normale Tastatur des Computers eingegeben, so besteht im Weiteren das Risiko, dass die Eingabe durch bösartige Software (Viren, etc.) mitprotokolliert wird. Auch der Einsatz eines Tokens nützt diesfalls nur wenig, weil diesem durch die bösartige Software später eine PIN-Eingabe vorgetäuscht werden kann, wodurch ohne Interaktion des Opfers eine Signatur erzeugt werden kann. Zur Erfüllung hoher Sicherheitsbedürfnisse sind derartige Lösungen daher ungeeignet. Um das Mitprotokollieren von Tastatureingaben zu verhindern, muss der Kartenleser oder das Token vielmehr unabhängig von der PC-Tastatur aktiviert werden können. Kartenleser etwa lassen sich dazu mit einer eigenen Tastatur ausstatten [1].

Biometrie
Durch die geschilderten, auf Besitz und Wissen basierenden Systeme kann nicht verhindert werden, dass der Inhaber des Tokens dieses zusammen mit dem Passwort weitergibt, oder dass dieses von einer Drittperson bei der Eingabe ausgespäht wird. Die Nichtabstreitbarkeit ist damit im Prinzip nicht gewährleistet.

Erst eine Prüfung anhand nicht übertragbarer, biometrischer Merkmale des Signierschlüsselhalters wie etwa Fingerabdruck, Unterschrift, Eigenarten der Stimme oder der Iris stellt sicher, dass wirklich nur der berechtigte Inhaber das Token einsetzt. Auch in diesem Fall muss natürlich die biometrische Prüfung durch das Token selbst oder zumindest durch ein vertrauenswürdiges Lesegerät durchgeführt werden, um zu verhindern, dass bösartige Software dem Token eine erfolgreiche Prüfung vorgaukelt. Tokens mit biometrischem Schutz des Signierschlüssels befinden sich derzeit in Entwicklung. Ein Problem bildet die Grösse der biometrischen Referenzdaten, die auf dem Token abzuspeichern sind; dieses wird mit der steigenden Leistungsfähigkeit der Chips jedoch einfacher lösbar. Weitere Bedenken bestehen hinsichtlich des Datenschutzes, der Anwenderakzeptanz sowie generell bezüglich der Sicherheit der Verfahren [1].

Anforderungen an Anwender und Software
Signierhard- und -software stellt regelmässig um so höhere Anforderungen an den Anwender, je einfacher der eingesetzte Mechanismus ist. So muss etwa der Anwender einer Softwaresigniereinheit mit Antiviren- und Firewall-Software dafür sorgen, dass sein System virenfrei bleibt und gegen Hackerangriffe abgeschirmt wird. Weil Computerlaien mit dieser Aufgabe regelmässig überfordert sind, muss deren Rechner als unsichere Zone gelten. Europäische Signaturgesetze verlangen daher vielfach den Einsatz von Hardwaresigniereinheiten mit vom Rechner unabhängiger Aktivierung [1].

3.4 Rechtliche Bestimmungen

> [...]they may not trust those systems with their business or personal interests unless there exists a suitable legal framework they can fall back on, should

problems arise[...] (Zitat von unbekannt)

Dieses Zitat von unbekannt spricht genau das Kernproblem an, dass kein System und keine Geschäftsbeziehung jemals wirklich vollständig so gestalltet werden kann, dass 100% Vertrauen gerechtfertig wäre. Daher soll eine rechtliche Grundlage genau da Ansetzen, wo die technischen und sozialen Aspekte zusätzliche Gewissheit fordern, damit eine Interaktion von Partnern via Web abgewickelt werden kann.

Rechtliche Erlasse oder Bestimmungen liegen in der Regel in der Authorität einer staatlichen Legislative. Besondesr da das World-Wide-Web aber ein globales Medium ist kommt sehr schnell die Frage auf, wie denn der rechtliche Rahmen überhaupt sinnvoll abgesteckt werden kann.

Im Moment ist es so, dass vorallem in vertraglich geregelten Wirtschaftsräumen Richtlinien und Erlasse wenigstens innerhalb eines geographisch beschränkten Wirtschaftsraumes halbswegs eine einheitliche Regelung für Ordnung sorgen sollen. Besonders die EU ist in diesem Bereich führend mit der EU-Signaturrichtlinie, E-Commerce-Richtlinie, die Fernabsatzrichtlinie oder die EU-Richtlinie 99/44/EG. Diese Richtlinien müssen dann von den Mitgliedsstaaten im nationalen Gesetzt umgesetzt und verankert werden.

3.4.1 EU-Signaturrichtlinie

EU-Signaturrichtlinie soll die Problematik der Signaturen bzw. der Zertifikate, die im Kapitel 3 angesprochen wurden, regeln.

Das Signaturgesetz ist am 22.5.2001 in Kraft getreten und spricht nicht von digitalen Signaturen sondern von elektronischen. Zudem wird zwischen zwei verschiedenen Formen von Signaturen unterschieden, der fortgeschrittenen elektronischen Signatur und der qualifizierten elektronischen Signatur. Die letztere basiert zum Zeitpunkt ihrer Erzeugung auf einem gültigen qualifizierten Zertifikat und muss mit einer sicheren Signaturerstellungseinheit erzeugt worden sein. Die fortgeschrittene Signatur soll gewährleisten, dass diese ausschliesslich dem Signatur-Inhaber zugeordnet ist, die Identifizierung des Signaturschlüsselinhabers ermöglich wird und dass die elektronische Signatur durch Mittel erzeugt worden ist, die aussschliesslich der Kontrolle des Inhabers unterliegen.

Auf nationaler Ebene wurde diese Richtlinie mit Signaturgesetzen umgesetzt und in der Regel durch eine unabhängige Kontrollinstantz oder Regulierungsbehörde in der Praxis realisiert. Eine solche Instanz oder Behörde stellt die Wurzelzertifikate aus, zertifiziert die öffentlichen Schlüssel der Zertifikationsstellen und überwacht auch diese Zertifikationsstellen [5]. Das Zusammenspiel sieht etwa folgendermassen aus:

3.4.2 Fernabsatzgesetz

Via Internet werden je länger je mehr Geschäfte auf dem Wege des Fernabsatzes abgeschlossen. Für den Internetnutzer resultiert daraus die Gefahr, die „Katze im Sack“ zu

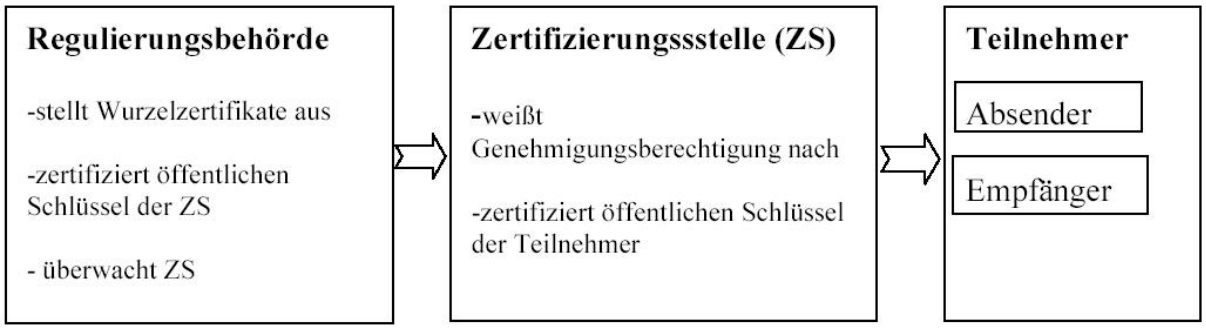

Abbildung 3.9: Übersicht Zusammenarbeit

kaufen. Denn zum einen kann der Kunde die Ware zuvor nicht in Augenschein nehmen und zum anderen kann er die Vertrauenswürdigkeit des Verkäufers im Vorfeld nur eingeschränkt überprüfen.

Aus diesem Notstand wurde auf EU-Ebene die Fernabsatz-Richtlinie erlassen. Die Fernabsatz-Richtlinie richtet sich an alle Geschäftsfälle, die über ein Fernkommunikationsmittel abgeschlossen werden (Telefon, Internet, etc.) und beinhaltet besondere Informationspflichten seitens des Anbieters und das Widerrufsrecht des Verbrauchers. Es ist besonders hervor zu heben, dass von diesen Richtlinien und der daraus resultierenden nationalen Gesetzen nur Unternehmen und private Verbraucher betroffen bzw. geschützt sind. Des Weiteren werden viele Geschäftsfelder auch wieder ausgeschlossen, insbesondere dann, wenn es dem Verkäufer nicht zumutbar ist, einem möglichen Widerrufsrecht seitens des Käufers ausgesetzt zu sein. Ebenso sind die meisten Tourismus-Dienstleistungen, Finanzgeschäfte und Versicherungsabschlüsse nicht davon betroffen.

3.4.3 E-Commerce-Richtlinie

Die so genannte E-Commerce-Richtlinie geht auf das Jahr 2000 zurück und hat es sich zum Ziel gesetzt, bestimmte Aspekte der Informationsgesellschaft zum Zwecke des einwandfreien Funktionierens des elektronischen Geschäftsverkehrs innerhalb des Binnenmarktes für alle für alle EU-Staaten verbindlich zu regeln.

Dazu gehören präzise und offene Angaben über den Anbieter, dessen Kontakt-Koordinaten und welcher Aufsichtsbehörde er allenfalls unterstellt ist, inklusive der Deklarierung von Registereinträgen wie Handelsregister oder ähnliche.

Des Weiteren muss offensichtlich erkenntlich sein, sobald es sich um kommerzielle Kommunikation handelt, die dahinter stehende Person in einer kommerziellen Kommunikation muss offenkundig sein, Sonderangebote, Verkaufsförderungsmassnahmen und Ähnliches muss als solches erkennbar sein und deren Bedingungen für die Inanspruchnahme müssen leicht zugänglich sowie klar und eindeutig angegeben werden und ebenso Preisausschreibungen oder Wettbewerbe.

Ein weiterer wesentlicher Bestandteil der E-Commerce-Richtlinie ist das Herkunftslandprinzip. Es besagt, dass für einen Anbieter jeweils die Gesetztgebung seines Herkunfts-

landes für ihn verbindlich ist. Das soll einerseits zu einer Vereinfachung im europäischen Online-Handel führen, andererseits besteht aber darin natürlich auch die Gefahr, dass lieberalere Gesetze eines Staates den Wettbewerb zu Ungunsten eines Online-Anbieters in einem stärker regulierten Staat verzerrt.

Das Herkunftslandprinzip wird jedoch zu Gunsten vom Verbraucherschutz wieder verwässert, so dass ein Kunde aus einem anderen europäischen Staat seine bessere Stellung vor dem Recht in seinem Land nicht verliert, wenn er ein Geschäft mit einem Online-Anbieter aus einem anderen Land abwickelt.

3.4.4 Rechtswahlklausel und anwendbares Recht

Was die Rechtswahlklausel und das anwendbare Recht angeht, besteht Wahlfreiheit für geschäftstreibende Parteien aus verschiedenen Nationen, sobald allerdings eine Partei davon ein Verbraucher ist, also in B2C-Beziehungen werden im Streitfall Einschränkungen durch das nationale Verbraucherschutzgesetzes aktiviert, damit Privatpersonen nicht des Schutzes ihres Heimatlandes beraubt werden.

Besonderheiten im B2C-Geschäft: Ein Online-Anbieter kann im Heimatland des Verbrauchers und in seinem Herkunftsland verklagt werden, der Verbraucher nur in seinem Heimatland.

UN-Kaufrecht: Eine weitere Besonderheit ist die Existenz des UN-Kaufrechts, das ein Übereinkommen darstellt über Verträge über den internationalen Warenkauf (CISG) das materielle nationale Recht verdrängt. Im Zweifelsfall kann das UN-Kaufrecht jedoch innerhalb von vertraglichen Bestimmungen ausgeschlossen werden, jedoch kommt es in der Regel zur Bestimmung zwischen Parteien, wenn davon mindestens eine in einem UN-Mitglied-Land wohnt.

3.4.5 Sonderfall AGBs / AGB-Richtlinie

AGBs werden in der Regel dafür verwendet Gewährleistungsrechte, Kündigungsfristen, der Umfang von Lizenzrechten oder Haftungsbegrenzungen bzw. -Freistellungen auszugestalten. Damit AGBs aber gültig sind muss der Kunde bis zum Vertragsabschluss auf ihre Geltung hingewiesen worden sein und er muss ohne Schwierigkeiten Zugang zu diesen haben.

Es gibt auf nationaler Stufe klare gesetzliche Grundlagen wie AGBs ausgestaltet sein dürfen und welche Klauseln rechtens sind. Sobald eine AGB-Klausel nationales Recht nicht einhaltet oder verletzt, muss die gesamte Klausel als ungültig angesehen werden.

Besonders bei B2C-Geschäftsaktivitäten muss der Anbieter in der Regel die Kenntnisnahme und das Einverständnis des Verbrauchers nachweisen, damit die AGBs bei einem Gerichtsfall Geltung hätten. Im Falle einer B2B-Angelegenheit reicht der Nachweis des deutlichen Hinweisens auf die AGBs. Zudem kann im B2C-Geschäft keine AGB den Verbraucherschutz verwässern.

Bei B2B-Geschäftsaktivitäten werden üblicherweise von allen involvierten Parteien AGBs formuliert. Dabei widersprechen sich einzelne Klauseln oft und darum werden jeweils so genannte Abwehrklauseln in die AGBs eingebettet, die entsprechende AGB-Klauseln der Gegenpartei für null und nichtig erklären sollen. Wenn nun beide Parteien solche Spielchen betreiben, gelten die betroffenen Klauseln beider Parteien für ungültig und es gilt das entsprechende nationale Gesetzt des Anbieters.

3.4.6 Internet- / IT-Standards

Neben der gesetzlichen Grundlage im Bereich des E-Business / E-Commerce werden in Zukunft vor allem IT- und Internet-Standards und -Zertifikate eine wichtige Rolle in der Vertrauensbildung spielen. Denn was nützt einem Verbraucher oder einem Businesskunden ein faires Angebot und ein zuverlässiger Anbieter, wenn seine persönlichen Bestellungsdaten ungeschützt auf einem Server des Anbieters liegen und von einem „Hacker“ dort gestohlen und anschliessend missbraucht werden [10]? In diesem Gefahrenbereich sind für Kunden aber auch für die Anbieter selbst noch verschiedene andere Szenarien denkbar. Daher könnte E-Businessteilnehmer ihre Vertrauenswürdigkeit durch IT-Audits, geprüfte Standards und andere qualitätssichernde Massnahmen zertifizieren und von Drittinstanzen beglaubigen lassen. Es müssten nicht einmal unbedingt staatliche Stellen sein, die solche Zertifikate ausstellen und die E-Businessteilnehmer prüfen, sondern NGOs oder Interessensverbände- und Konsortien können durchaus ebenso glaubwürdig sein und vor allem international einheitliche Standards erlassen [7]. Als Beispiel zu erwähnen wäre „Information Systems Audit and Control Association“, „Cobit“ und „ISO/IEC“ mit der Norm 17799, die in diesem Bereich schon weit fortgeschritten sind. Leider ist es so, dass

Abbildung 3.10: IT- Standart Zertifizierungsdienste

im Moment erst wenige Firmen sich selber um zertifizierbare Zuverlässigkeit und Qualität bemühen und viele Kunden bzw. Verbraucher die Anbieter nicht genügend unter Druck setzten mit einem Verlangen nach solchen qualitätssichernden Massnahmen [8]. Dabei könnte die gesamte Branche davon profitieren, wenn flächendeckende Standards und Massnahmen kommuniziert werden und umgesetzt würden.

3.5 Schlussgedanken

Trotzt allen Massnahmen und technischen Aufwendungen bleibt Vertrauen vor allem eine stark subjektiv bewertete Wahrnehmung oder Einschätzung von einer Situation. Und 100% Vertrauen wird es wohl nie geben, da keine 100% Zuverlässigkeit und Qualität in der Transaktion und im Geschäftsprozess gewährleistet werden kann.

Aber auch unsere direkte Wahrnehmung wird sich verändern und wir werden uns an neue Gepflogenheiten und Geschäftsmechanismen gewöhnen. Insbesondere aber werden sich auch die eingesetzten Technologien und Geschäftsprozesse laufend verändern und anpassen. Man bedenke, dass das Internet in seiner heutigen Form noch ein sehr junges Medium ist und wohl als Marktplattform noch keinen nennenswerten Reifegrad erreicht hat, sondern sich erst langsam zu etablieren beginnt und sich nach der „Dotcom- Krise“ um Nachhaltigkeit bemühen muss.

Schlussendlich wird sich in der kommenden Zeit auch die Rechtssprechung Schritt für Schritt auf die sich vollziehende Globalisierung einstellen müssen und dem neuen Medium „Internet“ mit einer moderneren Rechtsgrundlage gerecht werden. Erwähnenswert in diesem Bereich sind sicher, wie schon weiter oben aufgeführt, die relativ fortschrittlichen Richtlinien, welche von der EU in diesen Bereichen ausgearbeitet wurden.

Übergeordnete, internationale Legislativen werden neben der Rechtssprechung vermutlich in naher Zukunft noch andere Einigungen erzielen. Man könnte sich vorstellen, dass für verschiedene elektrische Verbrauchergeräte und andere Konsumgüter einheitliche Standards ausgearbeitet werden, die erfüllt sein müssen, damit sie angeboten werden dürfen. Oder internationale Standards für Medikamente, so dass wenn ich Heilmittel, Kosmetika oder Arzneimittel übers Internet einkaufe, einen gewissen Schutz und eine Gewähr habe, dass ich mich selbst nicht vergifte.

Viele von uns empfinden, betrachten und erleben das Internet zur Zeit als gesetzesfreien und nicht besonders vertrauenswürdigen Raum, man bedenke jedoch, dass auch der „wilde Westen“ zivilisiert wurde...

Literaturverzeichnis

[1] SIMON SCHLAURI Elektronische Signaturen DISSERTATION der Rechtswissenschaftlichen Fakultät der Universität Zürich S.47-98

[2] FRANZISKA ZUMSTEG Die Bedeutung von Vertrauen für den Erfolg von E-Government DIPLOMARBEIT im Fach Informatik angefertigt am Institut für Informatik 2004

[3] SANDY FRÄSSDORF Vertrauen als Bestandteil der Sicherheit im elektronischen Geschäftsverkehr DIPLOMARBEIT am Fachgebiet Informatik und Gesellschaft der Technischen Universität Berlin zur Erlangung des akademischen Grades „Diplom-Informatiker“ 2003

[4] FRANZISKA RÜEGG, ROGER SIGNER, ADRIAN HEYDECKER, NADIA VITALE Vertrauen und Glaubwürdigkeit im Online-Handel VORTRAG mit Notizen IPMZ Seminar „Glaubwürdigkeit in der Online Kommunikation“ 2004

[5] JÖRG BANGE, STEFAN MAAS, JULIA WASERT Recht im E-Business - Internetprojekte juristisch absichern 1. Auflage, Juli 2001

[6] ANDREAS DIEKMANN UND DAVID WYDER (2002): Vertrauen und Reputation bei Internet-Auktionen, Kölner Zeitschrift für Soziologie und Sozialpsychologie, 4/2002.

[7] THE ITGOVERNANCE INSTITUTE Cobit Framework

[8] CHRIS HARE Auditing the Electronic Commerce Environment

[9] INFORMATION SECURITY FORUM Securing E-Commerce (Workingpaper 1-5)

[10] RETO C. ZBINDEN Rechtliche Anforderungen, Standards und Best Practices (SWISS INFOSEC, 2. November 2004)

[11] EINWILLER S. (2003): Vertrauen durch Reputation im elektronischen Handel. Bamberg: Difo-Druck (Zugl.: St. Gallen, Univ., Diss., 2003).

[12] NICOLA DÖRING Sozialpsychologie des Internet, Die Bedeutung des Internet für Kommunikationsprozesse, Identitäten, soziale Beziehungen und Gruppen. Hoegrefe 1999.

[13] heise online - Neue Technologien unterminieren das Recht auf den Schutz der Privatsphäre, http://www.heise.de/tp/r4/artikel/1/1581/1.html. (Accessed: 21. February 2005)

[14] Digitale Signatur - Wikipedia, http://de.wikipedia.org/wiki/Digitale_Signatur, (Accessed: 10. February 2005)

[15] Sicherheit im elektronischen Geschäftsverkehr - Rechtsverbindlichkeit durch digitale Signaturen, http://ig.cs.tu-berlin.de/oldstatic/ap/rg/1998-06/abschnitt3.html, (Accessed: 10. February 2005)

[16] MasterCard - Switzerland German, http://www.mastercard.com/chd/, (Accessed: 10. February 2005)

[17] Visa Europe - Welcome to Visa, http://www.visaeurope.com/?country=ch&ep=v_gg_new, (Accessed: 10. February 2005)

[18] Zasterbox - Wirtschaftsauskünfte - Bonitätsprüfungen - Kreditmanagement / Forderungsmanagement, http://www.zasterbox.de, (Accessed: 10. February 2005)

Kapitel 4

Technology- and Layer-independent Authentication Schemes for the Internet

Andreas Wirth, Stefan Keller, Markus Stocker

Heute hat praktisch jede Internet-Zungangstechnologie wie PPP oder PPPoE eigene Authentifizierungs- und Authorisierungsverfahren. Durch eine Umschichtung der Authentifizierung von Layer 2 nach Layer 3 des ISO/OSI Modells, wäre es möglich, ein allgemeines Authentifizierungs- und Authorisierungsverfahren zu entwickeln, welches schliesslich unabhängig von der darunter liegenden Technologie funktionieren würde. Inhalt dieses Berichtes ist ein solches Verfahren welches aus einer technischen, praktischen und ökonomischen Perspektive genauer betrachtet wird.

Inhaltsverzeichnis

4.1 Einleitung

Die Verbreitung und Abdeckung kabelloser Netzwerke nimmt stetig zu. Zum einen ermöglichen uns die etablierten Mobilfunktechnologien GSM, GPRS und HSCSD beinahe weltweite, mobile Erreichbarkeit und weltweiter, mobiler Datenaustausch. Zum andern beginnt sich die neue Mobilfunktechnologie UMTS, zumindest in Europa, langsam auszubreiten. Ergänzt wird das Angebot mobiler Kommunikationsmöglichkeiten lokal durch WLAN Hot-Spots mit hoher Bandbreite. WiMAX könnte in Zukunft ebenfalls starke Verbreitung erlangen. Bereits werden erste Tests von Netzbetreibern durchgeführt.

Traditionelle mobile Geräte kommunizieren über einer dieser Technologien mit der Welt. Bereits sind aber Geräte erhältlich die sich mehrerer Kommunikationstechnologien bedienen. So kann ein PDA Benutzer per GPRS ins Internet gelangen oder sich innerhalb eines WLAN Hot-Spots mittels dieser Kommunikationstechnologie vernetzen. Die Wahl welche Technologie für die Kommunikation verwendet werden soll obliegt dem Benutzer. Dieser muss sich dann auch für jedes Netz separat authentifizieren. Die Übergabe einer Verbindung, beispielsweise vom GPRS in ein WLAN, ist nicht möglich. Das Schweizer Telekommunikationsunternehmen Swisscom hat kürzlich mit ihrer PCMCIA Karte 'Swisscom Unlimited' dieses Problem angepackt. Mittels dieser Karte gelangt der Benutzer dank WLAN, UMTS und GPRS ins Internet. Welche Technologie zum Einsatz kommt ist für ihn transparent und beim Wechsel von einer Netzwerktechnologie zur anderen wird die aktuelle Verbindung übergeben und reisst nicht ab.

In diesem Bericht geht es um technologie- und schichtunabhängige Authentifizierungs-Protokolle für das Internet.

Obwohl der Titel auf die Behandlung mehrerer Authentifizierungs-Schemas deutet, wird spezifisch auf das *Protocol for Carrying Authentication for Network Access* (PANA) eingegangen.

Das PANA Framework wird für ein Anwendungszenario, wie oben beschrieben, entwickelt. Die Möglichkeit einer Technologieunabhängigen Authentifikation und Authorisation ist Voraussetzung für das Übergeben von Verbindungszuständen über technologiefremde Netze hinaus. Die technologieunabhängige Authentifikation ermöglicht zudem ein für den Benutzer transparenter Wechsel der verwendeten Technologien - auch über ISP- und Netzbetreibergrenze hinweg.

Der Bericht wird die technischen Eigenschaften und eine Analyse des Einsatzes von PANA in heterogenen Netzen, mit Schwergewicht auf mobile Endgeräte, sowie PANA aus einer ökonomische Perspektive genauer betrachten und erläutern.

Auch werden Kritikpunkte dargestellt, meist bei Fragen zu Problemen für die PANA zur Zeit noch keine Antwort bereitstellt. Dazu gehört zum Beispiel die Problematik der Sessionübergabe bei der Verschiebung eines mobilen Gerätes von einer Antenne zur anderen oder vom einem Netz zum anderen.

PANA enthüllt sich schliesslich als noch sehr jung und bestimmt nicht Marktreif. Wie aber aus dem Bericht ersichtlich sein wird, hat das Protokoll auf jeden Fall ein gewisses Potential.

Ob PANA eine Zukunft im Massenmarkt hat, oder ob diese Technologie nur in Nischenprodukten zu finden sein wird, ist zum gegebenen Zeitpunkt sehr schwierig zu sagen.

4.2 Eigenschaften von PANA

Netzwerkzugangsauthentifizierung ist traditionell eine Funktion des Data Link Layers (Layer 2) des OSI-Modells [1]. So ist zum Beispiel *Point-to-Point* (PPP - Point-to-Point Protocol) ein Layer 2 Protokoll welches Teilnehmern, die keinem Netzwerk angehören (Dial-Up), den Zugang zum Internet ermöglicht [2]. Weitere Beispiele sind SLIP (Serial Line Internet Protocol) oder PPPoE (Point-to-Point Protocol over Ethernet).

Das Problem bei diesem Verfahren ist, dass kein allgemein übliches Protokoll für den Netzwerkzugang existiert. Zudem sollen die Zugangsdaten unabhängig von Zugriffsverfahren übermittelt werden. Ein weiterer Nachteil besteht in der Benutzung mehreren parallelen Zugangsverfahren. In diesem Fall müssen dementsprechend mehrere Verfahren parallel laufen, die grundsätzlich aber zum gleichen Ziel führen, nämlich die Netzwerkzugangsauthentifizierung eines Teilnehmers.

Bei PANA - Protocol for Carrying Authentication for Network Access [5] - handelt es sich hingegen um ein Netzwerkschicht-Protokoll (Layer 3). Diese Tatsache führt dazu, dass PANA unabhängig von der Zugangsart und der Netzwerktopologie ist.

Weiter ist PANA ein UDP - User Datagramm Protocol [3] - basiertes Protokoll[1].
Eine Anwendung die UDP nutzt, muss mit verloren gegangenen und umsortierten Paketen zurecht kommen, oder selber für entsprechende Korrekturmassnahmen aufkommen.

PANA geht den zweiten Ansatz indem ein eigener Übertragungs-Wiederholungsmechanismus implementiert wird, der die Nachrichten zuverlässig dem Kommunikationspartner übermittelt.
Dieser Mechanismus wird durch das Austauschen von Anfrage- und Antwortnachrichten zwischen den Kommunikationspartnern implementiert. Falls eine Partei eine Anfrage an die Gegenpartei sendet, letztere aber nicht antwortet, wird dies Protokolliert und die Anfrage kann bei Bedarf nochmals gesendet werden. Durch diesen Mechanismus kann überprüft werden, ob die Nachrichten korrekt ankommen.

PANA stützt sich auf EAP - Extensible Authentication Protocol [6] - für die Authentifizierung eines Clients.
EAP stellt ein grundlegendes Fundament für eine umfassende und zentralisierte Sicherheitskonzeption dar.
Es handelt sich um ein allgemeines Protokoll, das mehrere Authentifizierungsmöglichkeiten bietet. Es wurde ursprünglich für PPP entwickelt um eine zuverlässige Authentifizierung von Remote Access-Usern bereitzustellen [7].
Es werden in PANA also keine neuen Sicherheitsmechanismen entwickelt, sondern bestehende verwendet. Zur Authentifizierung bietet PANA lediglich das Übertragen von EAP Nachrichten zwischen dem Client und dem Server.

[1] UDP ist ein minimales, verbindungsloses Netzwerkprotokoll und ist nicht auf Zuverlässigkeit ausgelegt.

Bei PANA handelt es sich also um ein generelles Protokoll welches die Authentifizierung und Authorisierung eines Clients an ein Netzwerk als Aufgabe hat, wobei das Protokoll die Benutzerdaten, unabhängig vom Zugangsprotokoll und der darunterliegenden Infrastruktur, übermitteln kann.

4.3 Architektur von PANA

Das PANA Protokoll kommt zwischen einem Client (PaC - PANA Client) und einem Server (PAA - PANA Authentication Agent) zum Einsatz. Ziel ist die Authentifizierung und Authorisierung des Clients für den Netzwerkzugang.

Der Ablauf des Protokolls besteht aus einer Serie von Anfrage- und Antwort-Nachrichten. Diese Nachrichten enthalten verschiedene, sogenannte Attribute Value Pairs (AVP), welche es ermöglichen, relevante Informationen in Nachrichten zu kapseln und somit zu übermitteln.

Wie bereits der Name sagt, handelt es sich bei AVPs um Paare bestehend aus einem Attribut (Key) und einem Wert (Value). Dabei ist das Attribut festgelegt und gibt somit dem Wert der Nachricht eine Semantik. Die Bedeutung der Nachricht kann also durch das entsprechende Attribut vom Kommunikationspartner verstanden werden.

PANA Nachrichten werden als Teile einer PANA Session zwischen dem PaC und dem PAA gesendet. Solch eine PANA Session unterteilt sich, wie weiter unten noch gezeigt wird, in verschiedene Phasen.

Grundsätzlich muss als erstes der Client (PaC) einen Kommunikationspartner (PAA) suchen, bei welchem er sich Authentifizieren kann. Falls die Authentifizierung erfolgreich ist, wird dem PaC die Authorisierung zum Zugang ans Netzwerk erteilt.

PANA setzt sich aus den folgenden Teilen zusammen:

- **PANA Client (PaC)**
  Der PaC ist der clientseitige Teil des Protokolls, welcher sich im Host-Gerät befindet. Er ist verantwortlich für das Bereitstellen der Authentifizierungsparameter womit er seine Identität gegenüber dem Zugangs-Netzwerk geltend machen kann.

- **Device Identifier (DI)**
  Beim Device Identifier handelt es sich um eine Art Zeiger auf den Client. Dieser wird vom Netzwerk benutzt um den Zugang des Clients zu Überwachen und zu kontrollieren. Abhängig vom Zugangsverfahren kann sich der DI aus einer IP-Adresse oder auch ein vom lokalen Protokollstack zur Verfügung gestellten Identifizierer (wie zum Beispiel eine PPP Interface Id[2]) eines angeschlossenen Gerätes sein.

[2]Bei der PPP Interface Id handelt es sich um die Nummer des PPP Interfaces, meist 0 oder 1

- **PANA Authentication Agent (PAA)**
  Es handelt sich dabei um die Instanz des Zugangsnetzes welche für die Überprüfung der vom PANA Client (PaC) übermittelten Zugangsparameter verantwortlich ist (Authentifizierung). Ausserdem obliegt die Authorisierung [3] des über den Device Identifier (DI) identifizierten Gerätes in der Kompetenz des PAA.

Die Abbildung 4.1 soll graphisch darstellen, welcher Platz die jeweiligen Komponenten bei PANA einnehmen.

Es ist auch ersichtlich, dass ein Netzwerk unter Umständen mehrere PAA Server haben kann die auf Authentifizierungsversuche von aussenstehenden PaC's antworten können. Weiter unten wird erläutert, wie dieses Szenario gehandhabt wird.

Ausserdem kommt aus dem Bild hervor, dass der Device Identifier (DI) eine Eigenschaft des PaCs ist, welche diesen gegenüber dem PAA beim Aufbau einer PANA Session eindeutig identifizieren soll.

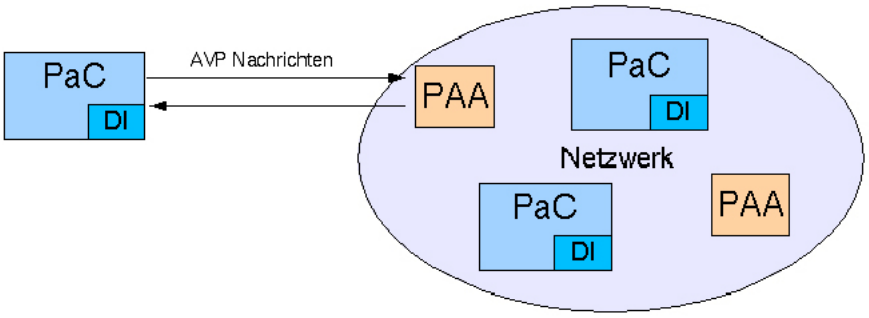

Abbildung 4.1: Beziehung zwischen einem PaC und dem für dessen Netzwerk-Authentifizierung verantwortlichen PAA. Stocker, 2004

4.4 Die Protokollphasen

Dieser Abschnitt erklärt im Detail die verschiedenen Phasen einer PANA Session.

Es gibt insgesamt fünf verschiedene Phasen. Eine davon (Re-Authentifizierung) setzt sich allerdings aus zwei anderen Phasen zusammen (Authentifizierung und Authorisierung). In diesem Sinne sind es nur vier verschiedene Phasen, mit dem Unterschied, dass die Authentifizierungsphase einer Re-Authentifizierung nur initialisiert werden kann, falls der PaC bereits eine existierende, gültige PANA-Session nachweisen kann.

Die Phasen folgen sich der Reihe nach, wobei jede Phase einen genauen Aufgaben- und Verantwortlichkeitsbereich hat.

Der Ablauf der Phasen, sowie die PANA-Nachrichten die zwischen dem PAA und dem PaC ausgetauscht werden, sind in der Abbildung 4.2 verdeutlicht.

[3] Das Vergeben der Rechte welche der Client im Netzwerk ausüben darf

```
PaC  PAA    Message[AVPs]
-----------------------------------------------------
// Discovery and handshake phase
 ----->     PANA-PAA-Discover
 <-----     PANA-Start-Request
 ----->     PANA-Start-Answer

// Authentication phase
 <-----     PANA-Auth-Request /* EAP Request */
 ----->     PANA-Auth-Answer
 ----->     PANA-Auth-Request /* EAP Response */
 <-----     PANA-Auth-Answer
 <-----     PANA-Bind-Request /* EAP Success */
 ----->     PANA-Bind-Answer

// Authorization phase (IP data traffic allowed)
 <-----     PANA-Ping-Request
 ----->     PANA-Ping-Answer

// Termination phase
 ----->     PANA-Termination-Request
 <-----     PANA-Termination-Answer
```

Abbildung 4.2: Übersicht der PANA Nachrichten innerhalb einer PANA Session [5].

4.4.1 Die Such- und Handshakephase

Wenn sich ein PaC an ein Netzwerk anschliesst, muss dieser ein PAA suchen. Bei

vorkonfigurierten Systeme ist die IP Adresse des PAA bekannt und die PANA-PAA-Discover Nachricht kann dieser Adresse direkt gesendet werden (unicast). Ist die Adresse nicht bekannt, so muss die PANA-PAA-Discover Nachricht an eine bekannte multicast Adresse und UDP Port geschickt werden, damit ein PAA gefunden werden kann. Dieser erste Schritt endet mit dem erfolgreichen Auffinden eines PAAs.

Der gefundene PAA bekommt also eine PANA-PAA-Discover Nachricht vom PaC. Auf diese Nachricht antwortet der PAA dem PaC (unicast) mit einer PANA-Start-Request Nachricht. Im Falle dass der PaC sich nun definitiv beim Netzwerk authentifizieren möchte, sendet er dem PAA darauf hin eine PANA-Start-Answer Nachricht.

Es ist natürlich möglich, dass mehr als nur ein PAA vorhanden ist und es kann ebenfalls sein, dass mehr als ein PAA auf die PANA-PAA-Discover Nachricht des PaC antwortet. Der Erfolg der Authentifizierung hängt nicht von der Wahl welchem PAA der PaC die PANA-Start-Answer Nachricht sendet ab. Standardmässig wählt der PaC den PAA, der ihm als erstes die PANA-Start-Request Nachricht sendet.

Die PANA-Start-Request Nachricht, welche der PAA als Antwort auf die PANA-PAA-Discover Nachricht dem PaC sendet, enthält ein Cookie welches DoS-Angriffe[4] auf den PAA verhindern soll. Das Cookie enthält einen zufällig generierten Wert der nur dem PAA bekannt ist (meist mit HMAC_SHA1[5] chiffriert).

In der PANA-Start Answer Nachricht des PaC befindet sich ebenfalls ein Cookie welches den vom PAA chiffrierten zufällig generierten Wert kapselt.

Das Cookie des PaC wird vom PAA auf Gültigkeit überprüft. Es ist gültig, wenn der Inhalt dem erwarteten Wert entspricht. In diesem Fall, geht das Protokoll in die Authentifizierungsphase ein, andernfalls wird die eingegangene PANA-Start-Answer ignoriert[6].

Es folgt ein Beispiel für den Aufbau eines solchen Cookies. Es gibt dafür keine exakte Beschreibung, da der Algorithmus zur Berechnung des Cookies die Interoperabilität nicht beeinflusst.

Cookie = <secret-version> | HMAC_SHA1(<Device-Id of PaC>, <secret>)

Dabei ist <secret> ein zufällig generierten geheimen Code der nur dem PAA bekannt ist. <secret-version> ist ein Index der für die Wahl des <secret> zur Erzeugung des Cookies benutzt wird. "|" bezeichnet die Verknüpfung.

<secret-version> sollte oft geändert werden, damit Attacken verhindert werden können. <secret> ist nur für eine gewisse Zeitspanne gültig. Die Device-Id des Clients kann aus einem Link-Layer oder IP Header der PANA Nachrichten entnommen werden.

Abbildung 4.3 zeigt die Sequenz der zwischen dem PAA und PaC ausgetauschten Nachrichten während der Such- und Handshakephase.

```
PaC      PAA          Message(seqno)[AVPs]
-----------------------------------------------------
   ----->             PANA-PAA-Discover(0)
   <-----             PANA-Start-Request(x)[Nonce, Cookie]
   ----->             PANA-Start-Answer(x)[Nonce, Cookie]
                      (continued to authentication phase)
```

Abbildung 4.3: Nachrichtenaustausch während der Such- und Handshakephase [5].

[4] http://en.wikipedia.org/wiki/Denial_of_service

[5] Keyed-hash message authentication code ist eine Art message authentication code (MAC) welcher für die Berechnung eine Kryptographische Hashfunktion kombiniert mit einem geheimen Schlüssel, benutzt [4]

[6] DoS-Angriffe werden somit verhindert, da der PAA die Authentifizierungsphase nur eingeht, wenn der PaC ihm korrekt antwortet. Der PaC kann also den PAA nicht mit Anfragen überfluten und dabei auf dem PAA grosse Ressourcen beanspruchen.

4.4.2 Die Authentifizierungsphase

Die Hauptaufgabe dieser Phase ist die Übermittlung von EAP Nachrichten zwischen dem PaC und dem PAA. EAP Anfrage- und Antwort-Nachrichten werden in sogenannten PANA-Auth-Request resp. PANA-Auth-Answer Nachrichten gekapselt übertragen. Letztere spielen meist keine grosse Rolle, denn PANA-Auth-Answer Nachrichten werden nur zur Bestätigung der Anfragen benutzt.

Die eigentliche Authentifizierung erfolgt durch das Extensible Authentification Protocol. Es liegt also an EAP zu entscheiden, ob die Authentifizierung erfolgreich ist oder nicht. Dieser Status wird durch den PAA über die PANA-Bind-Request Nachricht dem PaC mitgeteilt, auf die er mit der PANA-Bind-Answer Nachricht antwortet.

Diese beiden Nachrichten werden benutzt, um den Device Identifier (DI) des PaC und die IP des PAA mit der PANA Security Association zu binden. Bei der PANA SA handelt es sich um eine Verknüpfung zwischen dem PaC und dem PAA, die durch die gemeinsame Benutzung von kryptographischen Schlüsseln aufgebaut wird.

Durch die Chiffrierung der Daten zwischen dem PAA und dem PaC, wird einen Kommunikationskanal aufgebaut, der (nur) durch die entsprechenden Kommunikationspartner entschlüsselt und verstanden werden kann.

Abbildung 4.4 verdeutlicht den Nachrichtenaustausch zwischen dem PaC und dem PAA während der Authentifizierungsphase.

4.4.3 Die Authorisierungsphase

Sobald eine Authentifizierungsphase oder eine Re-Authentifizierungsphase erfolgreich beendet wurde, bekommt der PaC Zugang zum Netzwerk und kann IP Pakete senden und empfangen und die PANA Session geht in die Authorisierungsphase.

In dieser Phase werden sogenannte PANA-Ping-Request und PANA-Ping-Answer Nachrichten zur Überprüfung der Aufrechterhaltung der PANA Session benutzt (Abbildung 4.5). Sowohl der PaC als auch der PAA können PANA-Ping-Request Nachrichten dem Kommunikationspartner senden (Abbildung 4.6).

4.4.4 Die Re-Authentifizierungsphase

Eine PANA Session die sich in einer Authorisierungsphase befindet, kann eine erneute Authentifizierungsphase initiieren. Damit kann die Lebensdauer der aktuellen Session verlängert werden. Nach der erfolgreichen Re-Authentifizierung, startet die Session erneut eine Authorisierungsphase. Andernfalls wird die Session beendet.

Die PANA-Reauth-Request Nachricht ermöglicht dem PAA mitzuteilen, dass ein PaC eine Re-Authentifizierungsphase initiierend möchte. Diese Nachricht muss eine Session-Id beinhalten durch die die PANA Session auf dem PAA identifiziert werden kann. Falls der

```
PaC      PAA  Message(seqno)[AVPs]
-----------------------------------------------------
(continued from discovery and handshake phase)

   <-----     PANA-Auth-Request(x+1)
                 [Session-Id, EAP{Request}]
   ----->     PANA-Auth-Answer(x+1)
                 [Session-Id]
   ----->     PANA-Auth-Request(y)
                 [Session-Id, EAP{Response}]
   <-----     PANA-Auth-Answer(y)
                 [Session-Id]
   <-----     PANA-Auth-Request(x+2)
                 [Session-Id, EAP{Request}]
   ----->     PANA-Auth-Answer(x+2)
                 [Session-Id, EAP{Response}]
   <-----     PANA-Bind-Request(x+3)
                 [Session-Id, EAP{Success}, Device-Id,
                 IP-Address, Lifetime, Protection-Cap.,
                 PPAC, MAC]
   ----->     PANA-Bind-Answer(x+3)
                 [Session-Id, Device-Id, PPAC, MAC]
```

Abbildung 4.4: Beispielssequenz einer Authentifizierungsphase [5].

PAA die PANA Session identifizieren kann, sendet dieser dem PaC zuerst eine PANA-Reauth-Answer gefolgt von einer PANA-Auth-Request Nachricht, mit welcher eine neue EAP Authentifizierung gestartet wird. Wenn der PAA hingegen die PANA Session nicht identifizieren kann, sendet er dem PaC eine PANA-Error-Request Nachricht.

Abbildung 4.7 erläutert den Nachrichtenfluss zwischen dem PaC und dem PAA während einer Re-Authentifizierungsphase.

4.4.5 Die Beendigungsphase

Eine Anfrage zur expliziten Beendigung einer PANA Session kann sowohl durch den PaC als auch vom PAA imitierte werden. Dabei werden PANA-Termination-Request und PANA-Termination-Answer Nachrichten benutzt. Es wird ebenfalls der Beendigungsgrund dem Kommunikationspartner mitgeteilt.

Abbildung 4.8 verdeutlicht den Nachrichtenaustausch zwischen dem PaC und dem PAA während der Beendigungsphase.

```
PaC      PAA     Message(seqno)[AVPs]
-----------------------------------------------------
   ----->        PANA-Ping-Request(q)[Session-Id, MAC]
   <-----        PANA-Ping-Answer(q)[Session-Id, MAC]
```

Abbildung 4.5: Beispielssequenz für einen vom PaC initialisierten Ping-Test [5].

```
PaC      PAA     Message(seqno)[AVPs]
-----------------------------------------------------
   <-----        PANA-Ping-Request(p)[Session-Id, MAC]
   ----->        PANA-Ping-Answer(p)[Session-Id, MAC]
```

Abbildung 4.6: Beispielssequenz für einen vom PAA initialisierten Ping-Test [5].

4.5 Anwendungszenarien

Das im letzten Abschnitt beschriebene Authentifizierung- und Authorisierungsverfahren kann in den verschiedensten Szenarien zum Einsatz kommen. Zudem bietet PANA die Möglichkeit zur automatischen Auswahl eines Internet Service Providers (ISP). Durch die Kombination von PANA und traditionellen L2 Authentifikationsmechanismen kann eine Authentifikation unabhängig in verschiedenen Netzwerksegmenten durchgeführt werden. In den folgenden Unterabschnitten werden einige mögliche Anwendungszenarien genauer beschrieben.

4.5.1 PAA koexistent mit Access Router

In diesem Szenario koexistieren mehrere PAA nebeneinander. Dabei befinden PAA und Access Router (AR) auf der selben Maschine. Ein solches Szenario deckt beispielsweise den Fall ab, in dem mehrere Netzbetreiber verschiedene Funktechnologien bereitstellen. Wenn ein PaC authentifiziert und authorisiert werden muss (siehe Abbildung 4.9) tauscht er PANA messages mit dem PAA aus und eine Verbindung ins Internet wird aufgebaut. Bei einem Wechsel der Netzwerktechnologie erfolgt eine reauthentifizierung des Client durch den PAA. Bei einem Wechsel des Netzbetreibers würde der PAA des ersten Netzbetreibers die Authentifikations- und Verbindungsinformationen an den PAA des zweiten Netzbetreibers weitergeben.

In Szenarien, wie in Abbildung 4.9 dargestellt, stellt ein Netzbetreiber mehrere PAA/AR bereit. Beispielsweise könnte er die WLAN- und Ethernet-Verbindungen von einen, und die HSCSD- und UMTS-Verbindungen von einem anderen PAA authentifizieren und authorisieren lassen. Die Verbindungen werden dann auch über den entsprechenden Router geleitet um eine Verteilung der Last zu erreichen. Bei einem Wechsel der Netzwerktechnologie erfolgt dann die Übergabe der Authentifizierungs- und Verbindungsinformationen von einem PAA an den anderen PAA desselben Netzbetreibers.

```
PaC      PAA  Message(seqno)[AVPs]
-----------------------------------------------------
   ----->     PANA-Reauth-Request(q)
                 [Session-Id, MAC]
   <-----     PANA-Reauth-Answer(q)
                 [Session-Id, MAC]
   <-----     PANA-Auth-Request(p)
                 [Session-Id, EAP{Request}, MAC]
   ----->     PANA-Auth-Answer(p)
                 [Session-Id, MAC]
   ----->     PANA-Auth-Request(q+1)
                 [Session-Id, EAP{Response}, MAC]
   <-----     PANA-Auth-Answer(q+1)
                 [Session-Id, MAC]
   <-----     PANA-Auth-Request(p+1)
                 [Session-Id, EAP{Request}, MAC]
   ----->     PANA-Auth-Answer(p+1)
                 [Session-Id, EAP{Response}, MAC]
   <-----     PANA-Bind-Request(p+2)
                 [Session-Id, EAP{Success}, Device-Id,
                  IP-Address, Key-Id, Lifetime,
                  Protection-Cap., PPAC, MAC]
   ----->     PANA-Bind-Answer(p+2)
                 [Session-Id, Device-Id, Key-Id, PPAC, MAC]
```

Abbildung 4.7: Beispielssequenz einer Re-Authentifizierungsphase [5].

4.5.2 PAA separiert von Access Router

In diesem Anwendungszenario koexistiert der PAA nicht mit den Access Router, ist aber im gleichen Subnetz lokalisiert. Der PaC tauscht mit dem PAA die gleichen messages aus wie im letzten Szenario. Im Unterschied zum vorherigen Szenario erfolgt aber hier (Abbildung 4.10), nach erfolgreicher erster Authentifikation, eine Übergabe der Authentifikationsparameter an alle Access Router im Subnetz. Der PaC ist nun authorisiert seine Datenpakete durch alle Router in diesem Subnetz zu leiten. Ein solches Szenario währe in einem Fall denkbar in welchem ein Netzbetreiber sämtliche Netzwerktechnologien zur Verfügung stellt und der Benutzer mit diesem Betreiber einem Vertrag hat, der ihm erlaubt auch sämtliche Netzwerktechnologien zu benutzen.

Im Übrigen bietet das PANA Framework keinen Mechanismus zur Übergabe der Authentifizierungsparameter von einem PAA zu einem anderen. Dies ist nicht Bestandteil der PANA Spezifikation und wird daher auch künftig von einen separaten Protokoll erledigt werden müssen.

```
PaC      PAA     Message(seqno)[AVPs]
-----------------------------------------------------
   ----->        PANA-Termination-Request(q)[Session-Id, MAC]
   <-----        PANA-Termination-Answer(q)[Session-Id, MAC]
```

Abbildung 4.8: Beispielssequenz einer Beendigungsphase [5].

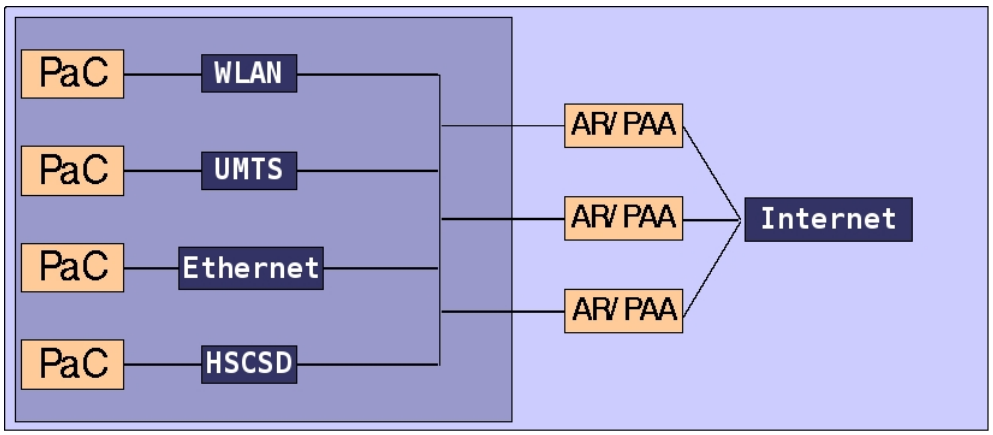

Abbildung 4.9: Koexistenz von PAA mit Access Router. Wirth, 2004

4.5.3 ISP Auswahl

Das PANA Framework sieht einen Mechanismus vor mit welchem einem Client automatisch ermöglicht wird sich mittels seines ISP mit dem Internet zu vernetzen. In Abbildung 4.11 entscheidet der PAA welcher ISP dem Client zur Verfügung steht und führt die Authentifikation und Authorisation durch. Beispielsweise könnte sich ein Benutzer in einem WLAN Hot-Spot befinden in welchem lokale Services angeboten werden. Damit sich der Benutzer mit dem Internet vernetzen kann muss er sich zuerst beim PAA authentifizieren. Der PAA wählt nun aufgrund der PaC-Identity den ISP automatisch aus und führt die Authorisation durch.

Dieser Mechanismus birgt allerdings ein nicht ganz triviales Problem. In einem solchen Hot-Spot bekommt ein Client typischerweise eine private IP-Adresse zugewiesen. Nach erfolgreicher Authentifikation/Authorisation und ISP Auswahl bekommt er nun eine IP-Adresse mit dem Address-Präfix seines ISP. Weitere Clients ddie, sich anderer ISP's bedienen, erhalten ebenfalls eine IP mit dem Präfix ihres ISP. In dieses Hot-Spot befinden sich nun Clients mit den unterschiedlichsten IP-Adressen. Weder das Routing innerhalb des Hot-Spots noch das Routing nach aussen ist nun sichergestellt.

4.5.4 L2 und L3 Authentifikation

Das PANA Framework funktioniert unabhängig von Authentifikationsmechanismen untenliegenden Schichten. Der Einsatz von PANA erlaubt also weiterhin eine Authentifikation auf der Sicherungsschicht(L2). Ein Betreiber eines WLAN Hot-Spots könnte beispiels-

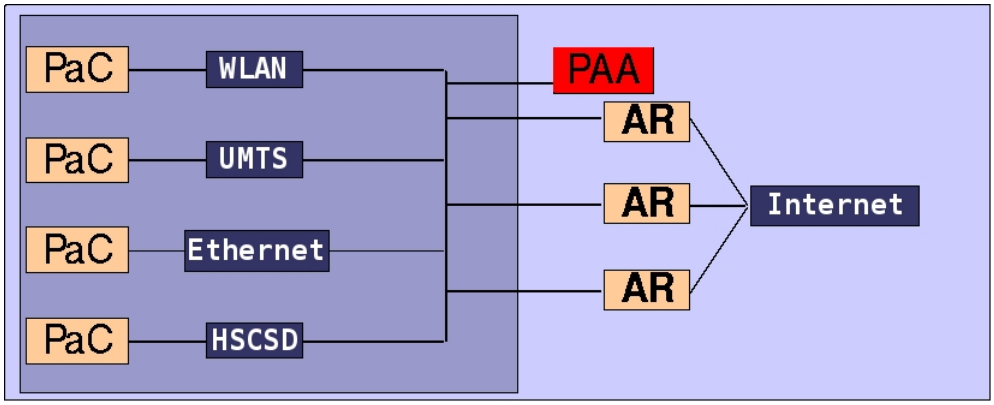

Abbildung 4.10: PAA separiert von Access Router. Wirth, 2004

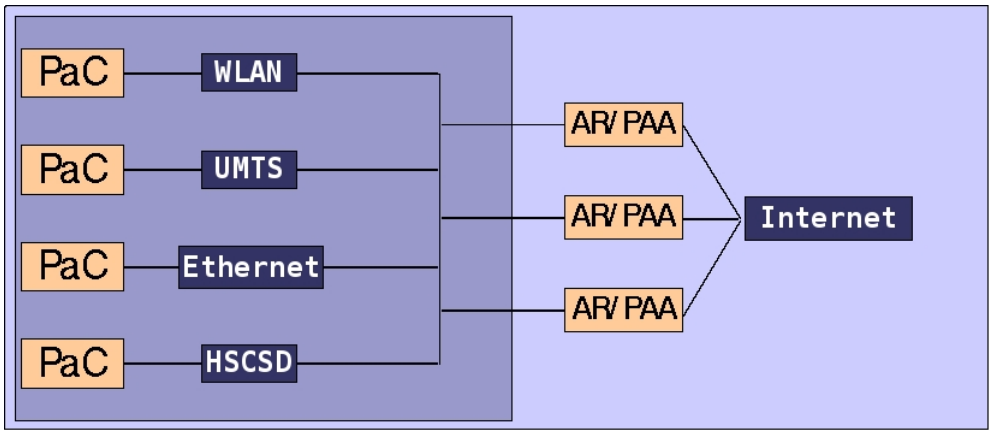

Abbildung 4.11: Automatische ISP Auswahl. Wirth, 2004

weise lokal die Benutzerauthentifikation und -authorisation auf L2 Ebene durchführen und diesen Benutzern dann Services anbieten (Abbildung 4.12). Möchte ein Benutzer ins Internet gelangen, so könnte er sich wie bisher gegenüber dem PAA authentifizieren.

Im oder obigen Abbildung würde sich der PaC zuerst im WLAN authentifizieren. Eine Möglichkeit währe, dies mittels eines L2-Tunneling-Protokolls wie z.B. (PPPoE) durchzuführen. Der Client kann nun den lokalen Service benutzen oder sich des Druckers bedienen. Der Betreiber des WLANs bietet hier aber keinen eigenen Anschluss ans Internet. Es ist Sache des Clients sich mittels PAA bei seinen eigenen ISP zu authentifizieren, welcher ihm dann den Zugang ins Internet ermöglicht.

4.5.5 DHCP

Die Authentifikation und Authorisation eines PaC setzt eine Gültige IP-Adresse für die Kommunikation zwischen PaC und PAA voraus. Wird in einem Subnetz die IP-Adresse per Dynamic Host Control Protocol (DHCP) vergeben, wird ein zusätzlicher Mechanismus benötigt um sicherzustellen dass nur authorisierte Benutzer auf das Netzwerk zugreifen können.

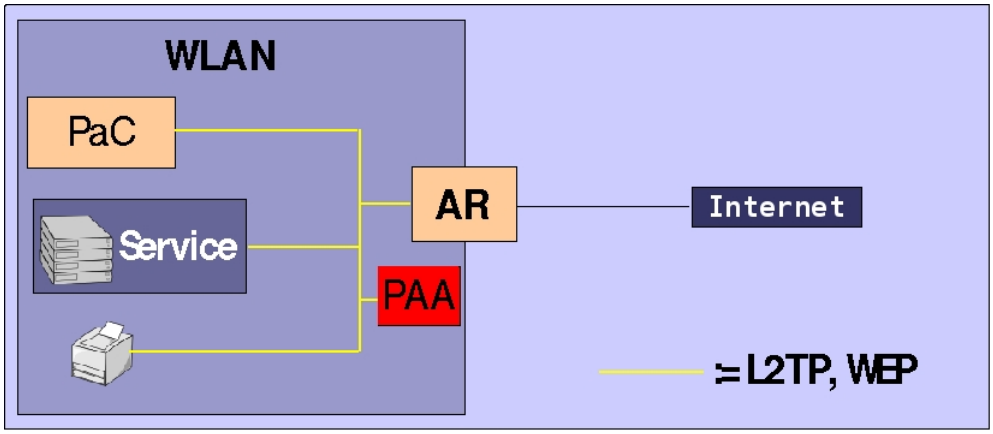

Abbildung 4.12: Authentifikation mittels L2-Protokoll und PANA. Wirth, 2004

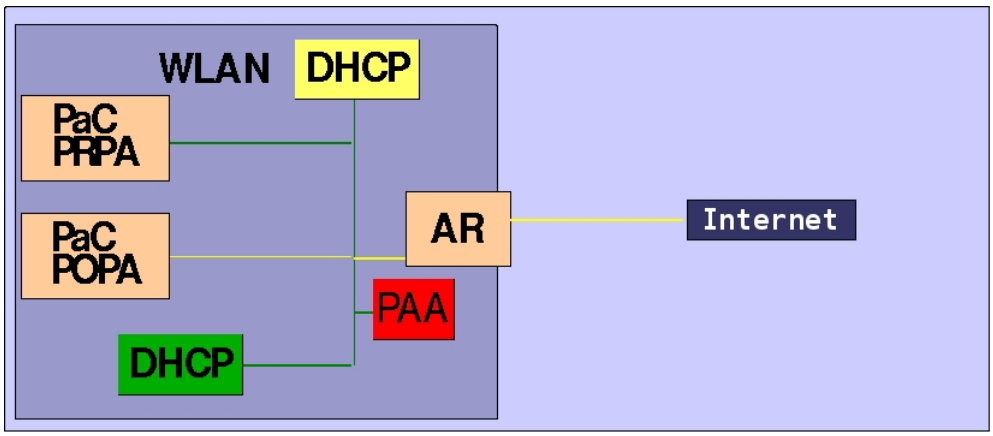

Abbildung 4.13: Authentifizierung für dynamische Adressvergabe. Wirth, 2004

Fordert ein PaC in einem WLAN eine IP-Adresse an, wird ihm vorerst von einem DHCP-Server (grün) eine temporäre Adresse zugewiesen. Das ist Beispielsweise eine Adresse die nur eine Zeitlich sehr begrenzte Gültigkeit hat, oder eine Adresse die nicht ins Internet geroutet wird. Nach Erhalt dieser temporären IP-Adresse ist der PaC nun fähig mit dem PAA eine PANA-Session auszuhandeln. Nach erfolgreicher Authentifikation und Authorisation kann sich der PaC von einem zweiten DHCP-Server (gelb), über den gesicherten Kanal, eine definitive IP-Adresse besorgen. Der PaC ist nun berechtigt das WLAN zu nutzen und über den AR ins Internet zu gelangen.

4.5.6 Mobilität

Der Benutzer eines Mobilen Gerätes, der Beispielsweise per Zug von einem WLAN in ein anderes reist, ist darauf angewiesen, dass währen des Überganges in ein anderes Netz seine TCP/IP-Verbindung nicht abreisst. Zudem währe es wünschenswert wenn beim Wechsel in ein WLAN, das eventuell von einem anderen Zugangsprovider angeboten wird, die PANA-Session in dieses neue Netz übergeben werden kann und keine erneute Authentifizierung stattfinden muss.

Die Übergabe einer PANA-Session ist also eine wesentliche Eigenschaft, die erfüllt sein muss, um den Benutzern einen transparenten Zugang zum Internet zu gewährleisten. Hier hat die PANA-Spezifikation bis zum heutigen Tag eine gewaltige Lücke. Diese Eigenschaft ist zwar in der Spezifikation aufgeführt, doch enthält sie bis anhin noch keinen konkreten Mechanismus der dies gewährleistet.

Die Übergabe der TCP-Verbindungszustände ist Voraussetzung dafür, dass während dem Wechsel von einem Netz in ein anderes die Verbindung nicht abreisst. Dies ist allerdings nicht Bestandteil der PANA-Spezifikation. Um einen nahtlosen Netzwechsel vollziehen zu können müsste ein Netztbetreiber, der PANA einsetzt, sich eines weiteren Protokolls bedienen. Dem dadurch entstehenden Flexibilitätsgewinn bei der Wahl des Handover-Verfahrens steht ein Anstieg der Komplexität der Lösung, ein grösserer Aufwand bei der Schulung des Wartungspersonals und möglicherweise eine erschwerte Wartbarkeit gegenüber.

Ob sich PANA durchsetzt hängt möglicherweise auch von den oben genannten Problemen ab. Sicherlich ist die Übergabe der PANA-Session beim Handover eine notwendige Bedingung. Ob dann die Vorteile eines offenen Standards ausreichen, um sich gegenüber proprietären Lösungen durchsetzen zu können, wird später in diesem Dokument aufgegriffen.

4.6 Der Markt für mobiles Arbeiten

In kabellose Netzwerke werden immer grössere Summen investiert. Alleine im Jahr 2004 wurden ca. 46.6 Millionen WLAN Units verkauft [10]. Gemäss Alexander Resources wird bis im Jahr 2007 ein Umsatz von 9.5 Milliarden Euros nur mit Public WLAN Spots generiert werden. 80% dieser Access Points wird in Cafés, Bars, Restaurants, Hotels, Flughäfen, usw. sein. Es ist jetzt aber schon klar dass der Grossteil dieses Umsatzes von Business Kunden generiert werden wird, und nicht von privaten Benutzern. Analysis Report spricht von ca. 80% Business Anteil. Anwender werden vor allem mit Laptops, PDAs, SmartPhones, usw. sich mit diesen WLAN Netzen verbinden.

Wie man heute schon sehen kann, bieten Firmen wie Swisscom oder IPASS WLAN Hot-Spots vor allem an Orten wo Geschäftsleute oft verkehren an. Dieser Trend sollte sich auch nicht sehr schnell ändern. Was sich sicher ändern wird, ist das man nicht mehr pro Datenmenge oder Stundenweise ein Zugangsticket lösen wird, sondern wie beim Mobilfunk eines oder mehrere Abonnements mit Flat-Rate bezieht.

Wie man in 4.14 sieht, werden Laptops immer verbreiteter. Jährlich reisen z.B. in den USA 43 Millionen Geschäftsleute, welche zu 70% einen Laptop haben, und zu 19% diesen für weniger als eine Stunde beim reisen benutzen. 31% benutzen diesen aber für mehr als 3 Stunden. Wie man aus jedem Laptop Prospekt sieht hat ein modernen Laptop WLAN schon beim Kauf integriert.

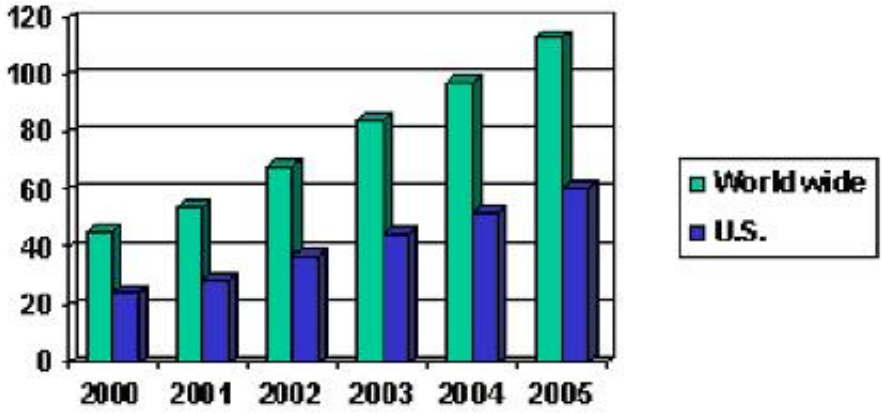

Abbildung 4.14: Anzahl Notebooks im Einsatz. Quelle: Cahner's In-Stat Group and Nokia 2001, Stand: April, 2001

Wenn wir analysieren was diese Business Kunden während dem mobilen Arbeiten machen, sehen wir, das alles eine Internet Verbindung benötigt oder sehr von Vorteil ist [11].

- E-mail lesen und schreiben 86%
- Dokumente bearbeiten 85%
- Im Internet surfen 74%
- Sich mit dem Firmen Netzwerk verbinden 59%
- Scheduling 48%
- Präsentationen vorbereiten 45%
- Entertainment 36%

Gemäss folgender Aussage könnte man meinen, PANA sei genau die Lösung dieser Probleme:

Business models must be formulated to support the multiple layers that are feeding the hot-spot market, from equipment vendors, to network providers, to roaming providers, to venue owners. Finally, marketing and branding efforts must be brought to the forefront as the amount and caliber of competition in this market escalates. [12]

Die verschiedenen Layer, Hardware und Software spezifischen AAA Methoden sind die grösste Herausforderung, um dem mobilen Benutzer Anschluss ans Internet zu bieten.

Wenn wir einen Blick auf die heute verfügbaren Netze (4.15) werfen, wird sich hier sicher auch nicht viel ändern, denn in den nächsten Jahren müssen vor allem die Mobilfunkanbieter ihre Milliarden schweren Investitionen in UMTS Lizenzen amortisieren.

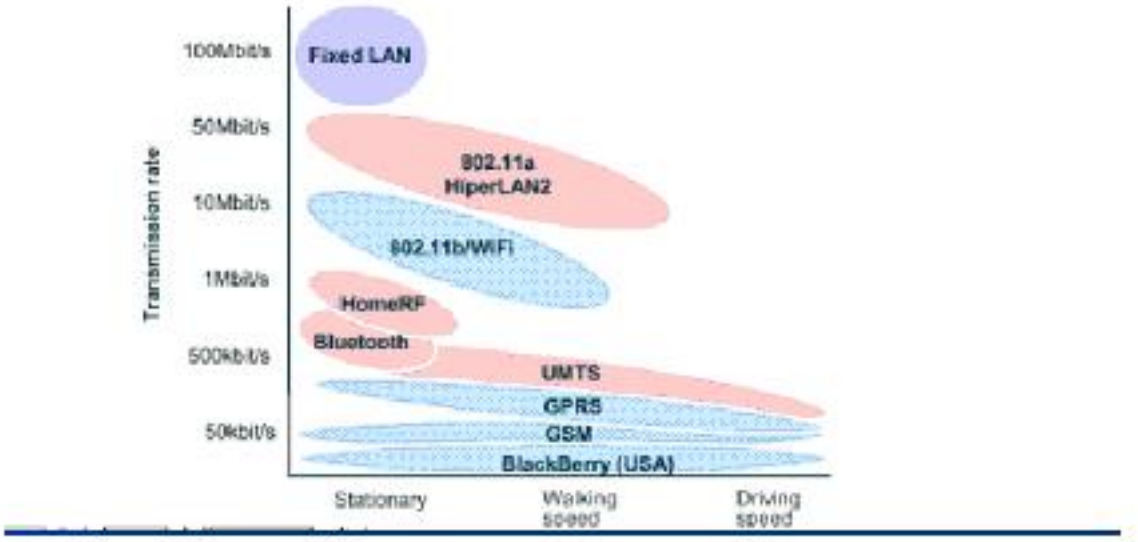

Abbildung 4.15: Verfügbare Funknetze, Quelle: Analysys Research

Das einzige was sich ändern könnte, ist das es eine flächendeckende Abdeckung mit WLAN Netzen geben wird, und man von einem WLAN Netzwerk in das andere mobil wechseln kann. Dies ist jedoch meiner Meinung nach sehr unwahrscheinlich, man bedenke nur wie viel Widerstand es gegen jede einzelne GSM Antenne in der Bevölkerung gibt.

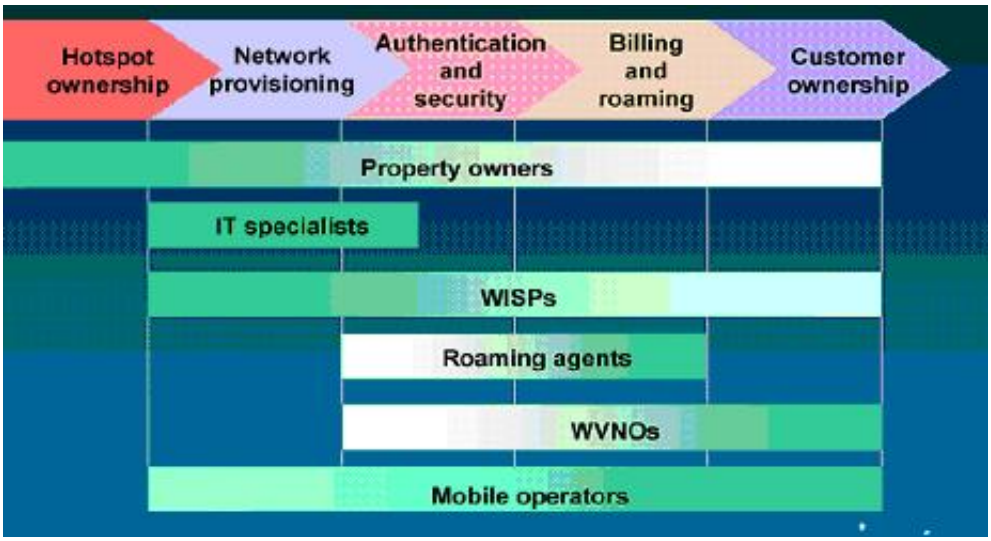

Abbildung 4.16: Mitbewerber welche an PANA interessiert sein könnten Quelle: 802.11 Planet, Stand: Juni 12, 2002

Wie wir in 4.16 sehen sind alle Anbieter von mobiler Kommunikation, von den in den vorherigen Kapiteln angebotenen Dienstleistungen, betroffen. Dies sind Hard- und Software-Entwickler, welche in den 3 tiefsten Schichten des ISO/OSI Modells Technologien Entwickeln, aber auch Wireless Internet Providers und klassische Mobilfunk Anbieter. Vor allem bei der Authentifizierung und Sicherheit, aber auch Rechnungsstellung und Ro-

aming, hat der PANA Standard viele Spezifikationen.

Aus kommerzieller Sicht haben sicher die heutigen Mobilfunk Anbieter den grössten Vorsprung, und sollten auch von einem offenen Standard wie PANA begeistert sein, da sie ihre aktuelle Position Nachhaltig sichern und ausbauen könnten. Dies scheint aber nicht der Fall zu sein, denn wenn man mit Google deren Webseiten durchsucht, oder auch die Mitglieder im IETF anschaut, sind da keine Mobilfunkanbieter vorhanden.

WISP können sehr schnell reagieren da sie Proprietäre Lösungen anbieten. Aber auch hier scheint das Interesse an PANA nicht vorhanden zu sein. Der einzige Anbieter welcher einen Satz über PANA sagt ist IPASS. Folgendes Statement von IPASS tönt jedoch noch nicht sehr überzeugend:

> *PANA addresses many of the requirements of public hot-spots and home broadband users; however the RFC is not completed at this time. [16]*

System Integratoren haben auch die Möglichkeit mit PANA einen offenen und unabhängigen Standard zu definieren, scheinen aber auch begrenztes Interesse daran zu haben.

4.7 Evaluation der Vorteilen und Nachteilen von PANA aus ökonomischer Perspektive

Hier wird anhand der folgenden Kriterien evaluiert, welches die Vor- und Nachteile von PANA sind:

- Technologie: ist PANA wirklich einmalig?
- Hat PANA eine Chance auf dem aktuellen Markt?
- Schafft es PANA eine fertige IETF Spezifikation zu werden?
- Wie lange dauert es noch bis PANA fertig spezifiziert ist?

4.7.1 Vorteile

PANA wird eine Link Layer unabhängige Authentifizierung ermöglichen Auf lange Sicht hinaus könnten neue Produkte auf der IP Ebene entwickelt werden, welche PANA benutzen können, und sich nicht um Link Layer spezifische Authorisierung kümmern zu müssen. Ohne PANA müssten für jede neue Link Layer Technologie, welche in den Markt eingeführt wird, jedes mal neue, Link Layer spezifische, Authentifizierungsmethoden entwickelt werden.

Hier findet man vor allem in der Betriebswirtschaftslehre das Netzwerk-Ökonomie-Modell, welches sagt, das eine Technologie oder ein Netzwerkgut wie PANA umso wertvoller ist,

je mehr Benutzer dieses schon haben. Klassische Herausforderungen sind hier die First Mover Vorteile (wer ist der erste im Markt? Wer macht wann welche Ankündigung?) aber auch Markt Lock-In Effekte, welche die Anbieter an eine Technologie binden und hohe Wechselkosten verursachen. Eine genaue Analyse ist momentan noch unmöglich denn PANA ist ja noch nicht fertig spezifiziert.

PANA ermöglicht das Roaming zwischen verschiedenen Netzen Da es eine IP basierte Technologie ist, wäre es möglich, wie im vorherigen Kapitel über Mobilität gesehen, zwischen verschiedenen Funknetzen und Technologien zu wechseln, ohne sich neu authentifizieren zu müssen.

Diese Ansätze scheinen mit UMTS, GPRS, und GSM sehr beliebt zu sein und könnten in Zukunft noch mehr Erfolg haben. Wenn eine Geschäftsfrau den Laptop einschaltet, und automatisch die schnellste und/oder billigste Internet Verbindung hat, und mobil arbeiten kann, wie im Büro als auch zu Hause, wird dieser Ansatz sicher sehr viel Erfolg haben.

IPv6 Mängel ergänzen Mobiles IPv4 ermöglicht das Authentifizieren mit einem unbekannten Verbindungsagenten, währen dessen das neue IPv6 diese Möglichkeit des Fremden Agenten nicht mehr bietet. PANA könnte diesen Mangel von IPv6 beheben und dies anbieten.

IETF Unterstützung Ausserdem liest man auf verschiedenen Quellen, dass PANA eine breite Unterstützung der Internet Area Directors der IETF geniesst. Es respektiert die bekannte Policy der IETF, die besagt, dass Link-Layer Funktionalität auf IP basierten Technologien auch von dem IP Protokoll behandelt werden sollte, und nicht von Link Layer Protokollen.

4.7.2 Nachteile

Viele Link Layer Technologien bieten schon Authentifizierungsmechanismen Die wichtigsten Link Layers, wie PPP und WLAN, besitzen schon gut implementierte Authentifizierungsmechanismen. Deshalb ist PANA nicht erste Priorität bei vielen Anbietern. Es ist nur auf lange Sicht aus interessant, falls neue Link Layer Technologien auf den Markt kommen sollten. Dies scheint aber in den nächsten Jahren nicht sehr wahrscheinlich zu sein, und bei der aktuellen Unsicherheit darüber, welches mobile Netz überhaupt wie viel Erfolg haben wird, auch keine Priorität zu geniessen. Wie schon in der Einleitung dieses Abschnittes gesagt müssen zuerst mal die Milliarden schweren Investitionen in das UMTS Netz amortisiert werden.

Ausserdem haben die wichtigsten Link Layers auch schon eigene Autorisierungsmechanismen, welche die Authentifizierung zuvor verlangen. Deshalb werden viele Anbieter von der Link Layer Mechanismen erst auf IP (und PANA) basierte wechseln, wenn die IP basierte

Authentifizierung möglich ist. PANA plant diese nachfolgende Authentifizierung erst später zu implementieren. Das dieser Wechsel nicht kurzfristig geschehen wird, unterstützt auch eine Aussage von J. Carlson aus der PANA Discussion List [17]:

> *Once again, I'm very confused about what it means to divorce authentication services from the fine-grained authorizations that these authentications often imply...*

Da sehr verbreitete Technologien wie 802.11b drei Kanäle haben, und die Möglichkeit Frequenzen mehrfach zu verwenden klein sind, besteht das Risiko dass Internet Service Providers anfangen sich WLAN Access Points sich zu teilen. In einem solchen Fall ist es wichtig dass verschiedene Benutzer autorisiert sind verschiedene ISP Netzwerke verwendet werden können.

Ein weiterer Nachteil ist, dass viele PANA Szenarien die Unterstützung von weiteren Protokollen erfordert, nicht nur von den IP basierten. Die Entwicklung und Unterstützung dieser zusätzlichen Protokolle benötigt sicher sehr viel Zeit und Geld. Nehmen wir z.B. ein Szenario wie es [13] beschreibt: Wenn der PANA Authentification Agent (PAA) nicht auf dem Access Point ist, dann muss ein neues Kontroll-Protokoll entwickelt werden (z.B. MIDCOM) damit der PAA den

Zugriff Verwalten kann. Dies würde zusätzliche Zeit benötigen um spezifiziert zu werden. Wenn der PAA hingegen auf dem Access Point sein muss, wird PANA die Multi-Hop Authentifizierung nicht unterstützen, was einer der stärksten Vorteile wäre, die eine IP basierte Authentifizierung ermöglichen würde.

Des weiteren liest man in [13] auch einige Kritik, dass PANA überhaupt nicht Link-Layer unabhängig sein soll:

- Aus Sicherheitsgründen wird die PANA Authentifizierung einen Geräte-Identifier enthalten. Hiermit wird z.B. die IP Authentifizierung mit der Link-Layer Adresse (oder Identität) verbunden. Dieser Device-Identifier wurde in der PANA List Discussion vorgeschlagen als man die User-Authentifizierung mit der Device-Authentifizierung verglichen hat. Deshalb folgt [13] zum Schluss das PANA nicht komplett Link-Layer transparent ist.
- Es ist sehr schwer sich vorzustellen, wie der Device-Identifier in existierende Authentifizierungsmechanismen eingebaut werden soll. Auf der anderen Seite, wenn die PANA Gruppe neue Authentifizierungsmechanismen entwickeln muss, wird dies sehr viel Zeit in Anspruch nehmen und eine sehr schwere Aufgabe sein.
- PANA könnte sich mit der schlechtesten Lösung zufrieden geben, und sich komplett auf die Link-Layer Authentifizierung stützen, aber dann wäre PANA ein sehr redundantes Protokoll.

Auch die Diskussionen über PANA in der Mailing Liste [18], bringen immer wieder neue technische Herausforderungen zum Vorschein, einige davon haben wir vorhin gesehen, und man muss die darunter liegenden Technologien auch berücksichtigen: Die Sicherheit, die Schnittstelle mit AAA, Zugangskontrollen, Device Identifier, usw. Auch wenn alle

diese Herausforderungen gelöst werden können, zeigt die Diskussion über PANA, dass die Entwicklung immer mehr Zeit benötigen wird als man am Anfang geplant hatte. Im Jahre 2002 glaubten viele Leute noch, dass man PANA entwickeln könne, indem man einfach die existierenden Authentifizierungsmechanismen in einen EAP Header einpackt, und dies mit UDP oder ICMP versendet. Deshalb glaubten viele, dass man Ende 2002 mit PANA fertig sein werde. Jetzt sind wird im Jahr 2005 und PANA ist noch nicht fertig spezifiziert. Dies sagt uns ganz klar dass die Lösung viel komplizierter ist, und es sehr schwer ist vorauszusagen, bis wann PANA fertig sein wird.

Es gibt auch einige einflussreiche Personen, welche gegen PANA sind. Sicherheits-Guru Glen Zorn, aber auch Pat Calhoun, ein Experte in Authentifizierungsmechanismen, sind Beispiele von Personen, welche ihr Widerstand gegen PANA kundtun.

Ausserdem liest man in [13], dass einige Mitglieder der PPP-Gruppe gegen PANA sind, da PANA eine direkte Konkurrenz zu PPP-basierten Mechanismen werden kann (not-invented-here Syndrom). Auch im IETF scheint es gemäss [13] einige Persönlichkeiten zu geben welche Widerstand ankündigen.

Zuletzt ist die Gruppe der aktiven PANA Mitglieder sehr klein, was die Entwicklungszeit natürlich länger macht. Auch Umfragen zeigen, dass nur wenige Hersteller aktiv sind. Man kann auch beobachten, dass die Aktivität auf der Mailing Liste seit dem Juni 2002 immer kleiner wurde.

4.8 Microsofts WLAN Strategie

Microsoft scheint an PANA sehr wenig Interesse zu zeigen, da sie ihre eigene Technologie, die Wireless Provisioning Services, bereits auf dem Markt lanciert hat. Dies ist eine neue Technologie für Public Hot-Spot Providers, welche automatisch die Netzwerkselektion und das Sing-Up erledigt.

Die Microsoft Wireless Provisioning Services (WPS) Technologie ermöglicht einem Netzwerk Provider, eine auf existierenden Standards basierte Wi-Fi Technologie, und eine integrierte Plattform um die Wi-Fi Spots zu verwalten [19]. WPS ermöglicht es Windows XP Benutzern sich bei diesen Public WLAN Hot-Spots sehr bequem und einfach anzumelden. Da dieser Dienst auf der existierenden Wireless LAN Technologie basiert und bekannte Tools wie der Canncection Wizard, Sicherheits Features wie die Protected Extensible Authentification (PEAP) und Wi-Fi Protected Access (WPA) benutzt, sind keine neuen Technologien oder Protokolle nötig.

Auf der Provider Seite ist ein Windows 2003 Server nötig, der schon alle Komponenten enthält. Das vereinfacht die Installieren, die Konfiguration und die Verwaltung deutlich [20]. Ausserdem basiert diese Technologie auf dem Internet Authentification Service (IAS), welcher auch Microsoft RADIUS Server genannt wird, und dank der schon bestehenden Integration auch die Total Cost of Ownership, gegenüber neuen und nicht verbreiteten Technologien wie PANA, massiv reduziert. Microsoft Lösung basiert hier auch auf nicht Proprietären Standards, auch für Virtual Private Networks nicht.

Ausserdem ist diese Lösung für Nutzer von Windows XP und Windows 2003 Server kostenlos als Zusatz erhältlich.

4.8.1 Komponenten der Mircosoft Wireless Provisioning Services

Auf der Microsoft Webseite findet man sehr konkrete Anleitungen darüber, was genau die Komponenten sind, um diese Public WLAN Hot-Spots einzurichten (4.17). Dies ist bei PANA nirgends so klar und deutlich aufgezeichnet, was einen Kunden sicher verwirrt und die Entscheidung für die Microsoft Technologie erleichtert.

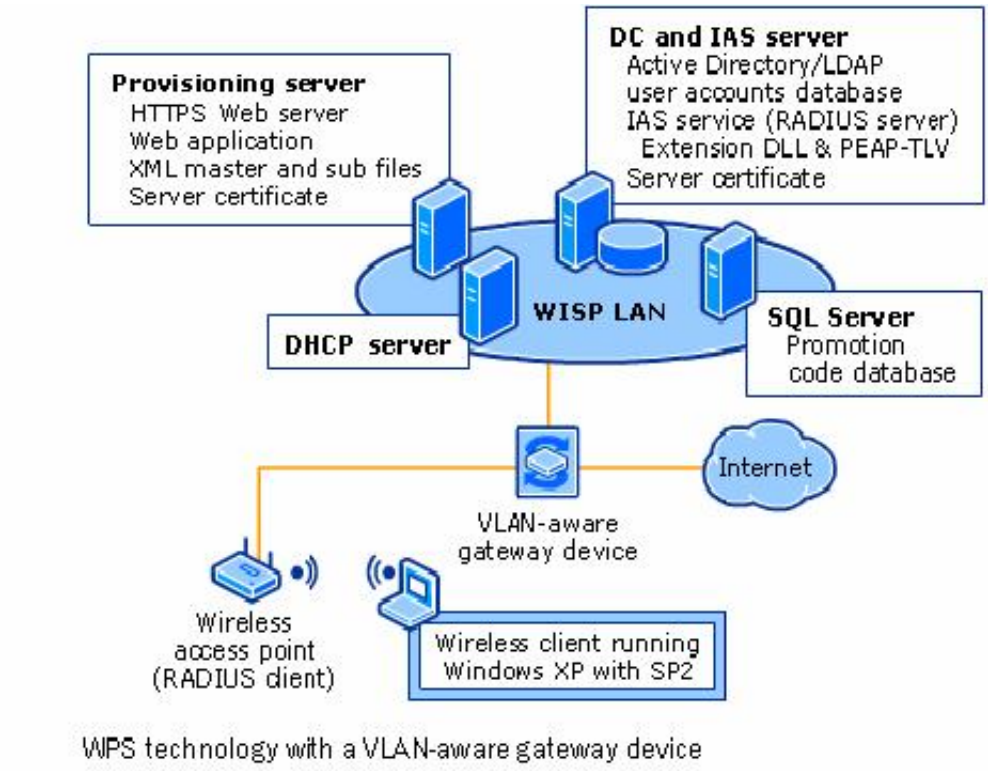

Abbildung 4.17: Microsofts WPS Technologie Architekturbeispiel

Die folgenden Komponenten sind ein Beispiel wie eine Microsoft Lösung aussehen könnte:

- WPS kompatible wireless Geräte
- Wireless access points (APs) mit der Unterstützung von virtual local area networks (VLANs) oder Internet Protocol (IP) filtering
- Zugangskontrolle (Access controller)
- Provisioning server
- Active Directory® Service als Domain Controller wo alle User, Geräte, Drucker, Rollen, Berechtigungen, usw. aufgelistet sind und verwaltet werden

- Internet Authentication Service (IAS) server für die Authentifizierung
- Dynamic Host Configuration Protocol (DHCP) server, damit z.B. den neu angemeldeten Clients dynamische IP Nummer zugewiesen werden können.

Überblick der Komponenten der Microsoft Lösung

4.8.2 Aussagen aus der IT Industrie über die neue Microsoft Technologie

Es ist sehr leicht viele Statements der Industrie zu finden, welche die neue Microsoft Technologie unterstützen, über PANA ist auf den gängigen Suchmaschinen, aber auch in der Literatur, nur sehr schwer positive Aussagen zu finden.

Die nachfolgenden Aussagen bestätigen die breite Unterstützung der Microsoft Lösung. [14]:

„Boingo is a pioneer in bringing Wi-Fi hot-spots to the masses, so we're delighted to be working closely with Microsoft on its Wireless Provisioning Service initiative. Microsoft is helping the industry to standardize on key protocols that will make Boingo's job of aggregating the world's largest hot-spot roaming system a lot easier."

Niels Jonker, CTO, Boingo Wireless Inc.

„Cometa Networks supports the development of the Wireless Provisioning Service (WPS) architecture. As we deploy the Cometa hot-spot network, we are committed to drive down customer care costs through simplification of the end-user experience. WPS helps create a more seamless, transparent and secure user experience that will help drive adoption of broadband wireless services."

Jeff Damir, Senior Vice President, Cometa Networks

„GRIC believes Microsoft's Wireless Provisioning Service provides a common framework to help users to discover and connect between wireless hot-spots in an easy and secure fashion. WPS removes the ease-of-use and security obstacles to widespread Wi-Fi adoption by enterprise end users."

Lumin Yen, Director of Wireless and Broadband Product Management, GRIC Communications Inc.

„The standards-based, vendor-neutral access client that Microsoft is developing should greatly increase takeup of Wi-Fi. For example, through our Wi-Fi Roaming Services offering, iPass can enable carriers and mobile operators to leverage our Enterprise Ready global network of Wi-Fi hot-spots, while giving carriers the power and flexibility of using access technology with the most widely deployed operating system in the world."

Roy Albert, Vice President of Product Development, iPass Inc.

„Microsoft Wireless Provisioning Services technology provides a consistent end-user experience and security features that have been lacking so far in the public Wi-Fi market. Our Operations Support System for public hot-spots supports Microsoft's standards-based WPS technology, and we anticipate that our service provider customers will benefit greatly from the automatic and seamless provisioning and configuration capabilities it provides."

Jasbir Singh, President and CEO, Pronto Networks Inc.

„As a leading provider of Wi-Fi service in hotels and airports nationwide, Wayport views Microsoft's Wireless Provisioning Services technology as a well-thought-out methodology that allows customers to easily subscribe to and use Wi-Fi service."

Dan Lowden, Vice President of Marketing, Wayport Inc

Ausserdem Arbeitet Microsoft auch sehr aktiv mit dem IETF und den IEEE Standard bodies zusammen, aber auch mit Herstellern wie Agere, Cisco und Enterasys. Aktuell ist der Fokuss auf der Verbreitung und Akzeptanz des IEEE 802.11i Standards.

4.9 Herausforderungen für das Roaming

In einer sehr ausführlichen Intel Studie [15] werden die grössten Herausforderungen für eine das Public WLAN Roaming sehr fein und klar dargestellt.

- Jeder Operator/Carrier hat sein eigenes Business Modell und unabhängige Standards, welche PWLAN Roaming und Interworking ermöglichen.
- Hot-Spots Installationen werden nicht von einem einzigen Operator oder einer Operator Community betrieben, weil die kleine Reichweite, der tiefe Preis der Hardware und die fehlenden Regulationen viele dazu motiviert ihre eigene Infrastruktur aufzubauen. Aber die Benutzer werden sicher nicht ein Account für jeden einzelnen Zugriff eröffnen.
- Der grosse Erfolg der Mobilfunk Netze ist eine grosse Konkurrenz für PANA und PWLAN.
- Die Technologie, und dessen schnelle Entwicklung, fördert eher eine Markt Fragmentierung anstatt eine lokale und globale Interoperabilität.
- Die Akzeptanz bei den Benutzern wird langsam wachsen, da viele Dienstleistungen nicht kosteneffizient, überall verfügbar und sicher sind. Ausserdem ist die Installation und Konfiguration oft sehr schwer.

Intel beschreibt den PWLAN Roaming Prozess in 4.18 mit den folgenden fünf Schritten:

Wie man in Abbildung 4.18 sieht, wird nachdem ein Mobile Device das Netzwerk entdeckt und ausgewählt hat (Schritt 1.), eine gegenseitige Authentifizierung vorgenommen und die Verbindung hergestellt (Schritt 2.). Schritt 3. ist sehr wichtig, da man hier differenzierte

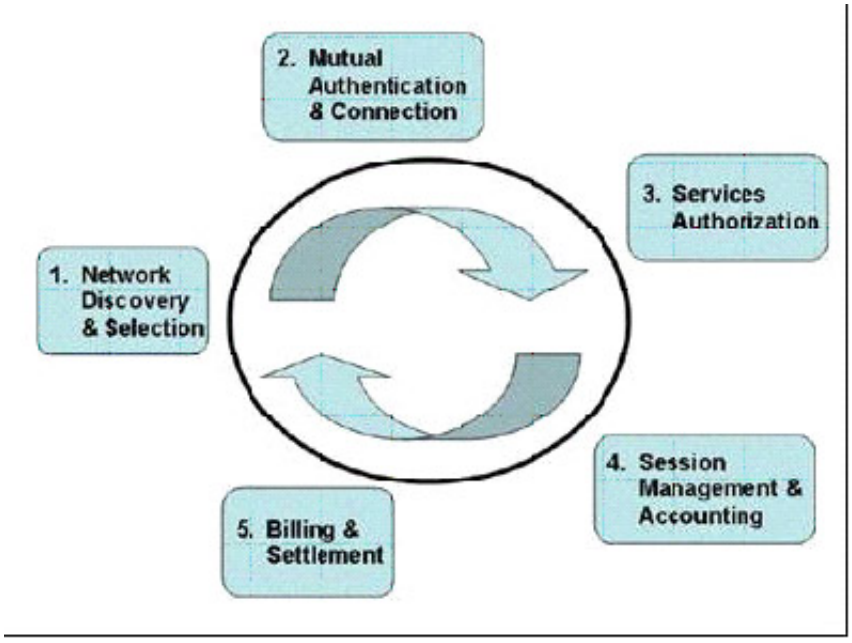

Abbildung 4.18: Fünf Schritte für das PWLAN Roaming

Services anbieten kann, z.B. aufgrund der bezahlten Summe, der Stufe, usw. Nach der Autorisierung muss die Session Verwaltet und gezählt werden (Schritt 4.). Am Schluss muss der Provider dem Home Provider des Users eine Rechnung stellen (Schritt 5.). PANA muss sich aus ökonomischer Perspektive diesem Standard anpassen, damit es auch mit den Business Modellen der Provider kompatibel ist.

Wie man in 4.19 sieht, müssen die AAA Prozesse bei dem Home Provider des Kunden erfolgen, dies bringt aber auch folgende Risiken und Herausforderungen bei einem praktischen Einsatz von PANA für das PWLAN Roaming:

- End-To-End Security muss gewährleistet sein. Wie man in Schritt vier und sieben sieht, erfolgt dies über das Internet. Hier müssen Verschlüsselungsmechanismen und weitere Sicherheitsvorkehrungen getroffen werden.

- Privacy: Roaming Partner sollte den User nicht erkennen, muss ihn aber trotzdem Authentifizieren wie man in Abbildung 4.20 ebenfalls sieht.

- Risiko des Kunden Diebstahls durch den Roaming Anbieter falls dieser zu viele Informationen über den User hat.

- Service Levels müssen auch im Roaming differenziert werden: einem User der mehr bezahlt sollten erweiterte Dienstleistungen oder mehr Services angeboten werden als einem der weniger bezahlt. Die Information welcher Zugriffslevel dem mobilen User gewährt werden, müssen im vierten Schritt zwischen den beiden AAA-Servern übermittelt werden.

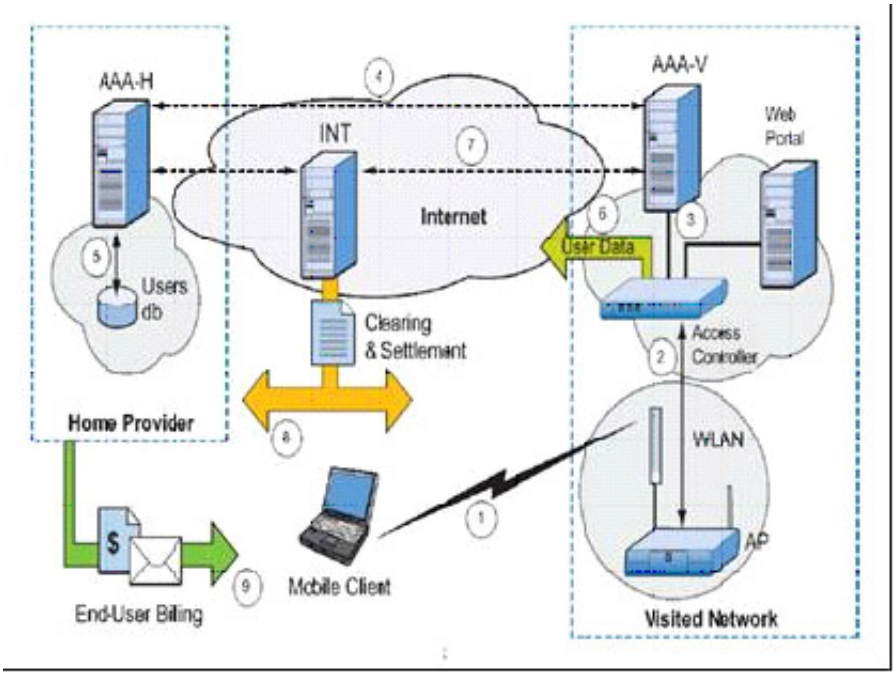

Abbildung 4.19: Generelles Roaming Modell

4.10 Alternative mobile Geräte für PWLAN Benutzung

Auf dem Markt gibt es schon einige Geräte welche mobiles E-mail, WWW, Telnet, usw. anbieten. Diese Hardware hat zwar direkt nichts mit PANA zu tun, könnte aber ein Risiko für PANA darstellen, denn diese Geräte ermöglichen es den Benutzern ebenfalls via Mobilfunkanbieter (Roaming) sich mit GPRS, UMTS, oder wie in den USA mit Verizon, mit PWLAN Netzten zu verbinden. Anfangs Januar 2005 war einer der Autoren in New York City in einem Hotel beim Times Square und hatte im 35. Stock perfekten Empfang des Verizon WLAN Netzwerk via BlackBerry.

Die oben vorgestellte Swisscom Unlimited Lösung ist auch schon existent und im Einsatz. Es ist deshalb sehr fraglich ob PANA überhaupt noch benötigt wird, und ob es eine Möglichkeit hat vom Markt akzeptiert zu werden, wenn das Mobile Arbeiten mit folgenden Geräten so einfach ist:

- BlackBerry:

  Der BlackBerry ist ein Handheld und Mobilfunktelefon mit GPRS Internet Verbindung, 30 Mio. Users Weltweit, Real Time E-mail, Integration von Microsoft Exchange Kalender mit Wireless Syschronisation, E-mail Attachments, WWW Browser.

  Der SPV M2000 ist ein auf Windows Mobile basierter Handheld mit Mobilfunktelefon, integrierter WLAN Karte, Internet Explorer, Bluetooth, E-mail Client mit IMAP und POP Support, GPRS.

Abbildung 4.20: BlackBerry von RIM und das SPV M2000 von Orange

4.11 Schlussfolgerungen

Die PANA Vision ist gut, und könnte nützlich werden, wenn man das Roaming zwischen verschiedenen Netzwerk Technologien ermöglichen will.

Aber es wird sicher noch einige Zeit brauchen bis PANA einsatzbereit ist, da wie wir gesehen haben, einige technische Probleme noch gelöst werden müssen, und eine starke Konkurrenz zu anderen Lösungen herrscht.

Bedenklich ist das Microsoft sich überhaupt nicht für PANA interessiert, und eine fertige Lösung schon entwickelt, und auf dem Markt gebracht hat. Diese ist zwar nicht so offen und standardisiert wie PANA das sein wird, aber sicher schon viel verbreiteter.

Ausserdem birgt das WLAN Roaming noch sehr viele Probleme bei den sehr heterogenen Business Modellen der verschiedenen PWLAN Anbietern. Diese müssten sich für einen effizienten PANA Einsatz alle einigen und einen einheitlichen Service anbieten.

Zuletzt ist es fraglich ob man PANA dann überhaupt noch benötigt wenn immer mehr mobile Geräte wie der BlackBerry auf dem Markt kommen, welche schon alle Funktionen des Internets mobil anbieten.

Literaturverzeichnis

[1] http://en.wikipedia.org/wiki/OSI_Model

[2] http://www.ietf.org/rfc/rfc1134.txt

[3] http://www.ietf.org/rfc/rfc0768.txt

[4] http://en.wikipedia.org/wiki/HMAC

[5] http://www.ietf.org/html.charters/pana-charter.html

[6] http://www.ietf.org/rfc/rfc2284.txt

[7] http://de.wikipedia.org/wiki/Extensible_Authentication_Protocol

[8] http://www.informatik.fh-nuenberg.de /professors/schwenk/SII/NetworkAccess/PANA_vortrag_mehrer%20final.pdf

[9] http://sourcewire.com/releases/rel_display.php?relid=20096&hilite=

[10] http://www.instat.com/catalog/nqcatalogue.asp?id=160#IN0401426WL

[11] Cahner's In-Stat Group, Stand: June, 2000

[12] In-Stat/MDR reports, Stand: June, 2002

[13] http://www.unik.no/ paalee/publications/Telenor-PANA-note-2002.pdf

[14] http://www.microsoft.com/presspass/press/2003/Oct03/10-12WPS2003PR.asp

[15] http://www.intel.com/technology/roaming/irap/

[16] http://www.ipass.com/pdfs/GIS_whitepaper.pdf

[17] http://www.atm.tut.fi/list-archive/pana/msg01029.html

[18] http://www.atm.tut.fi/list-archive/pana/msg00896.html

[19] http://www.microsoft.com/presspass/press/2003/Oct03/10-12WPS2003PR.asp

[20] http://www.microsoft.com/windowsserver2003/technologies/networking/wifi/default.mspx

Chapter 5

DDoS and other malicious Attacks - Effects on Enterprises' Budgets

Alexander Müller, Lukas Keller, Nicolas Lüscher

One of today's great challenges of modern technology is computer network security and especially preventing malicious attacks from targeting certain entities. When a company that uses the Internet for delivering its services or products is facing such an attack, big money is at stake. The last few years showed us that not even large companies that spend millions of dollars in infrastructure are not 100% prepared to face such an attack. In this chapter an overview of the most frequent types of attacks that are used nowadays for inflicting damage in a company's network infrastructure shall be presented. A closer look at DDoS attacks and similar threats that have happened in the past shall be taken, their effects discussed and their trends examined. In a further step we look at current solutions that the market offers for protecting from such attacks and try to venture a view into the future. Based on those technical considerations and details investigated, the following key economic question shall be answered: What is cheaper: to prevent or to repair?

Contents

5.1 Introduction

Many critics have described the current era as the information age, the dawn of a bright new future, a time when the barriers to communication have been by-passed, allowing the formation of virtual communities all over the globe. Businesses now have the ability to expand their presence beyond the normal limits of geography, enabling them to reach out to a market that years earlier they would have, by necessity, ignored. Users all over the Internet share information and experiences almost instantly with people a hundred of miles away. The application of Internet technology and the associated opportunities seem endless. And that is part of the problem. With every opportunity there comes risk. In the world of the Internet, this risk often materializes in the form of security. Internet and security are inevitably connected: one should always accompany the other. When using the Internet, security should always be an important subject. However, some even believe, that already using the Internet excludes the fact, that one can be completely secure. There is still much research to be done and probably the good guys will always be one step behind the bad. The different security threads, which an administrator comes in contact with during his work, sound like words out of a science fiction book: highly contagious viruses, infiltrating trojans, autonomous living worms or even attacking zombies.

Since 1999 a new threat exists, expanding the list of dangers, one is confronted, when connected to the Internet: the Distributed Denial of Service (DDoS). All around the globe security experts had predicted this sort of menace, hoping it would never come true. New tools for performing Denial of Service (DoS) attacks on a massive scale were released to the Internet. These new tools were referred to as DDoS tools, because of their distributed nature. They allowed an attacker to coordinate attacks against Internet sites from client machines, mostly called zombies, distributed around the world, using a single client program. Given enough zombies, an attacker could enforce any site to its knees.

Figure 5.1: Bringing civilization to its knees...

As the security community realised the dangers, these tools created, they tried to alert the rest of the world. But it was too late, the assaults already had started. In just a few short days, the foundations of some of the largest Internet sites were rocked by massive coordinated attacks. Because bandwidth had become a commodity, with broadband access offering high-speed Internet connectivity through cable modems and asymmetrical digital

subscriber lines, conditions were already available, which made it possible for the thread, to assume such alarming proportions. Unaware of the dangers the Internet-users faced, the penetrations began occurring at an alarming rate, leaving behind massive networks of DDoS zombies for later use. Another sad fact was that many of the largest sites on the Internet not even had basic protection mechanisms implemented and therefore were no match for the zombie-hordes from a DDoS-attack. The combination of technological advancement and the circumstance of being unaware of the new thread allowed a single David to knock down several Goliaths with one powerful stone (see figure 5.1).

Before the reader is confronted in the first part with the concrete composition and functionality of a denial of service attack, we shall first introduce and explain the general existing malwares, like viruses, worms and trojan horses. The latter shall be compared with an illicit server. Hereon we shall take a look at the most common attacking mechanisms a hacker can apply. Afterwards we lay the groundwork for understanding DDoS: we present the DoS attack. For that, we present two elemental types of DoS attacks: the resource consumption attack (such as SYN flood attacks and amplification attacks) and the malformed packet attack. Finally we conclude the first part with analyzing the denial of service attack.

In the second part we analyze first of all the present situation and define the actual threats of today. We discuss the types of electronic crimes and examine the identity of the attackers. In a next part, we show you the possible preventive and reactive countermeasures against DDoS attacks in relation to the different target groups. We try to answer questions like, which countermeasures can be suggested as the most important and useful ones, what are possible costs, which can occur and so on.

In the third and last part of this chapter we dare to take a look into the future and try to speculate about the possible trends. We shall analyze law developments, eventually new or changing customer interests and directions, in which the hacker techniques progress. Afterwards we present you possible new worms, which could eventually arise in the near future. New malware, with much more intelligence, updateable functions or even polymorphic behaviour! In e next part we discuss the computer security trends of tomorrow, examine Microsoft's philosophy of trustworthy computing and look at the distribution of operating systems and web browsers. Finally possible financial consequences for enterprises shall be examined and a brief conclusion shall be made.

5.2 The taxonomy: viruses, worms and trojan horses

5.2.1 Viruses

A virus is any binary file that meets the following criteria:

1. It requires direct human intervention in order to spread. Unlike a worm, which spreads automatically, a virus requires a user to download and double-click a binary file, or transfer it using an infected medium, such as a floppy disk.

2. It has a payload, which can be destructive behaviour (deleting or altering files), or annoying messages left on the screen, or both.

3. A virus spreads quickly to all documents in an operating system. A virus never spreads itself to other systems automatically.

Although many different types of viruses exist, like boot-sector-viruses, file-viruses, polymorphic viruses, retro-viruses, script viruses, resident viruses, stealth viruses etc., the macro viruses are by far the most common. Word processors and spreadsheets, such as Microsoft Word and Excel, allow users to create powerful, convenient mini-applications that reside within the word processor. These macros are meant to simplify life by cutting down on repetitive tasks. The problem with macros is that many end-users allow macros to run without first establishing controls over what they can do. The macro facilities in office suites, such as MS Office, are almost always powerful enough to launch applications, delete files, and begin a sequence of events that can seriously damage the system [3]. A malicious user can take advantage of powerful macro facilities. In fact, the Melissa virus is a macro virus. Many others exist that are not as ambitious, but which are still powerful.

5.2.2 Worms

The chief difference between a worm and a virus is that a worm spreads to other systems. Furthermore, a worm is able to spread with little or no user intervention. Remember, in order for a virus to spread, a user must first install it by copying a file or inserting a floppy disk. A worm can spread itself upon activation. By simply double-clicking a file, the worm can be activated, and deliver its payload (if any), then spread by taking advantage of system settings, macros, and applications (called application programming interfaces, or APIs) that reside on a system. Whereas a virus is generally designed to spread throughout an entire machine, a worm is designed to propagate itself to all systems on a network [3]. There are four factors that allow a worm to spread rapidly:

1. Networks that use one operating system. For example, an exclusively Microsoft or Novell network stands a greater risk of rapid infection than a heterogeneous network that uses UNIX, Novell, and Microsoft servers.

2. Networks that standardize to one mail user agent (MUA), such as Microsoft Outlook. Just as networks that have one operating system are vulnerable, a company that uses one MUA is liable to experience an event where a virus is propagated quickly. Also, because Outlook is so popular, hackers are more familiar with it. Therefore, a hacker can create an application that exploits it.

3. Operating systems, such as those vended by Microsoft, that provide interpreters and models, such as the Component Object Model (COM), which make it easy to create powerful applications in just a few steps.

4. Networks that use TCP/IP. Although TCP/IP is a powerful, efficient protocol, it was not designed with security in mind. Although the next version of IP, called

IPv6, improves security, this version of IP has not been implemented widely. The current version of IP, called IPv4 allows a malicious user to imitate respectively spoof the origin of an IP address. As a result, it can be very difficult to find the true attacker in case of an incident.

Types of worms

Below is a brief discussion of the three major types of worms:

1. **True worms**: Requires no human intervention to spread. This type of worm is rare, because it requires great skill on the part of the programmer, and will function only on a homogeneous network. A true worm is also rare because it uses the programming language of the e-mail server itself. For example, to create a worm for the Netscape Enterprise e-mail server, you would have to write the application using the language that Netscape Enterprise Server uses.
2. **Protocol worms**: Any worm that uses a transport protocol, such as TCP/IP, to spread. The Robert Morris worm, for example, used elements of TCP/IP, including finger and Sendmail (which uses SMTP), to spread itself. This type of worm can also spread without any direct human intervention.
3. **Hybrid worms**: A worm that requires a low level of user intervention to spread, but also acts like a virus. A simple click on a malicious attachment does not mean that this user is ready to copy or transmit an application. However, a click still represents user intervention. Most of the worms, like BubbleBoy, Melissa, and Life Stages are hybrid worms, because they behave like viruses in that they deliver a payload. However, they also exhibit worm-like behaviour, because they are able to spread automatically from system to system.

5.2.3 Trojan horses

A trojan horse, or trojan, is nothing more than an application that purports to do one thing, but in fact does another. Trojans are named after the mythic trojan horse in Homer's Iliad. In the legend, the Greeks created a wooden horse and gave it to the citizens of Troy as a peace offering. However, before the horse was presented, Greek soldiers hid inside it. The horse was brought inside the city gates, and when the city was asleep, the Greek soldiers emerged and were able to conquer Troy. Similarly, a trojan looks like a useful program, but contains a payload. For example, a trojan can:

- Launch an application that defeats standard authentication procedures.
- Delete files.
- Format the hard drive.
- Launch legitimate applications with the intent of defeating security.

Many Trojans have a payload. A common payload is to delete a file, many files, or even an entire partition. Perhaps the most common payload is an illicit server [3].

Illicit servers

An illicit server is nothing more than a simple service or daemon that defeats a server's authentication mechanisms. A valid server, such as an e-mail or web server, always has authentication mechanisms that allow only certain users. Illicit servers have the following characteristics:

1. They open up a TCP or UDP port (over 1024) often only for a short time.
2. They attempt to hide any trace of their existence. They do not show up in a task bar or in a task list.
3. Most of the time, an illicit server is a very small binary that is easy to conceal as a hidden file, or it is one small file in the midst of several others.

Using such a server, a malicious user can compromise an e-mail server. Examples of illicit servers include:

- **NetBus and NetBus Professional**: Although many professionals consider NetBus Professional to be perfectly legitimate, each of these applications can be used to gain unauthorized control of a system. NetBus has a client and a server. Usually, a hacker will engage in social engineering or other means in order to get the server installed on the victim's system.
- **Back Orifice and Back Orifice 2000**: More ambitious than NetBus, these illicit servers allow you to open FTP and HTTP connections on any port you specify. Using these servers, a malicious user can read the entire hard drive of any Windows system, as well as upload, download, and delete files. Back Orifice 2000 even allows a malicious user to specify a password, encrypt transmissions, and even destroy the server to avoid detection. Like NetBus, Back Orifice uses a client and a server.
- **Netcat**: Although a legitimate tool, it is possible for a malicious user to use this application to create an illicit server.

Many other illicit servers exist, most of which one will never hear about; after all, why would a hacker give up trade secrets? Usually, a hacker will trojanize these servers in an attempt to trick end-users into installing them. Such social engineering practices are common. One of the more infamous examples of social engineering is where a hacker took a version of the Whack-A-Mole game and linked it to NetBus. Then, the hacker began sending this game to various people, who then played it and unwittingly installed the NetBus server on their systems.

Differentiating between trojans and illicit servers

One can not use the terms trojan and illicit server interchangeably. An illicit server is often presented to users in trojanized form, but an illicit server is not necessarily a trojan. For example, unless one disguises NetBus as another application, it is simply an illicit server [3].

5.2.4 Hoaxes

A hoax is actually no malware in terms of a virus, trojan, etc., but its consequences are not to be underestimated. Briefly said, a hoax is an e-mail containing an announcement over malware, which does not exist. Almost all those false messages describe new, never before seen and very dangerous viruses or worms, which have outstanding destructive consequences. Hoaxes mostly are written in an exaggerated manor, trying to initiate the panic and fear of a reader, using well known sources like Microsoft, McAfee, Symantec, etc., to convince and proof the trueness of the message. At the end of the message follows normally the advice to forward the e-mail to all family members, relatives and friends, that they also get warned over the new threat. This assignment is actually the ”replication engine” of this pseudo-virus. In forwarding the hoax, a pyramid scheme gets started and the ”virus” spreads faster than any other real malware. It’s like someone shouts ”fire” on a busy, crowded net and the e-mail equivalent of a stampede begins. That’s in the end also the purpose of the hoax: to frighten the innocent user and to stress mail-servers and the Internet itself.

5.3 Attack mechanisms and principals

One shouldn’t make the mistake of thinking that hackers simply attack systems. Many different types of attacks exist. Some require more knowledge than others, and it is often necessary to conduct one type of attack before conducting another. In the following a list of the common attacks waged against all network-addressable servers is presented [3].

5.3.1 Scanning

Most of the time, hackers do not know the nature of the network they wish to compromise or attack. By using TCP/IP programs such as ping, traceroute, and netstat, a hacker can learn about the physical makeup (topology) of a network. Once a hacker knows more about the machines, it is possible to attack or compromise them.

5.3.2 Denial of service attacks

This type of attack usually results in a crashed server. As a result, the server is no longer capable of offering services. Therefore the attack denies these services to the public. Many of the attacks started against e-mail servers have been denial of service attacks. However, one should not confuse a DoS attack with other attacks that try to gather information or obtain authentication information. A more detailed report over DoS attacks and its extended version, the DDoS attacks follow later on.

5.3.3 Sniffing and/or man-in-the-middle attacks

This attack captures information as it flows between a client and a server. TCP/IP is an inherently insecure protocol, because it does not encrypt transmissions by default. Therefore, it is possible for a malicious user to use a protocol analyzer (also called a packet sniffer) to capture and then view packets. Applications such as Sniffer Basic and TCPdump are specially designed to place a Network Interface Card (NIC) into promiscuous mode. Once in promiscuous mode, a NIC can then capture any packets that are passing through your particular portion of the network. Usually, a hacker attempts to capture TCP/IP transmissions, because they may contain information such as user names, passwords, or the actual contents of an e-mail message. A sniffing attack is often classified as a man-in-the-middle attack; because in order to capture packets from a user, the machine capturing packets must lie in between the two systems that are communicating (a man-in-the-middle attack can also be executed on one of the two systems).

5.3.4 Hijacking and/or man-in-the-middle attacks

Another form of a man-in-the-middle attack is where a malicious third party is able to actually take over a connection as it is being made between two users. Supposing, that a malicious user wants to gain access to machine A, which is beginning a connection with machine B, the malicious user creates a denial of service attack against machine B. Once the hacker knocks machine B off of the network, he or she can then assume that machine's identity and collect information from machine A.

5.3.5 Physical attacks

So far, all attacks introduced, are started from one remote system to another. But it is also possible to walk up to the machine and log in. For example, how many times one simply just walks away from a machine after having logged in? A clever hacker may be just waiting for a careless user, who lets his computer unattended in an office or a public place, to take over his system and assume his identity. Other, more sophisticated, attacks involve using specialized floppy disks and other tools meant to defeat authentication.

5.3.6 System bugs and/or back doors

No operating system, daemon, or client is perfect. Hackers usually maintain large databases of software that have problems that lead to system compromise. A system bug attack takes advantage of such attacks. A back door attack involves taking advantage of an undocumented subroutine or (if one is lucky) a password left behind by the creator of the application. Most back doors remain unknown. However, when they are discovered, they can lead to serious compromises.

5.3.7 Social engineering

The motto of a good social engineer is: Why do all the work when one can get someone else to do it for oneself? Social engineering is a computer-expression for the practice of leading someone into revealing too much information. Many social engineers are good at impersonating system-administrators. Another example of social engineering is the temporary agency that is, in reality, a group of highly skilled hackers who infiltrate companies in order to conduct industrial espionage.

5.3.8 E-mail bombing

Another form of attack involves sending hundreds, if not thousands, of large e-mail messages to an account on a server. Due to the large volume of e-mail messages (not to mention their size), the victim account will remain unusable until the systems administrator removes all of the messages, or creates another account. Many easy-to-use applications exist that are meant to enable the most untalented user to send an e-mail bomb.

5.3.9 Spamming

Many older message transfer agent (MTA) servers allow any user or system to connect to them and send e-mail anonymously. Whenever an e-mail server allows a user to send e-mail anonymously, it is said to allow relaying. Servers that allow relaying allow users to specify any user name and any DNS domain in an e-mail message. For example, should you find an e-mail server that allows relaying, you could, with just a few commands, create a fairly convincing e-mail message from bill.gates@microsoft.com, william.shakespeare@bard.com, or georg.bush@whitehouse.org. While this practice may seem amusing, bulk e-mail applications can send thousands, if not millions, of junk e-mail messages called spam. Although most MTA servers that currently ship do not have relaying turned on, one can still find systems which have it enabled. Not only is spam e-mail annoying, it wastes time, valuable network bandwidth and slows down the Internet. The Mail Abuse Prevention System (MAPS) is one of several organizations that have organized to prevent spamming (www.mail-abuse.org). Their chief goal is to conduct scans of e-mail servers across the Internet and then inform systems administrators that their servers currently allow e-mails to be sent anonymously. MAPS then informs the offending system administrator. If no

action is taken, then MAPS will blacklist the e-mail server so that it cannot communicate with the rest of the Internet.

5.4 What is a DDoS attack?

To understand a DDoS attack and its consequences, one first needs to learn the fundamentals of DoS attacks. The progression from understanding DoS to DDoS is quite elementary, though the distinction between the two is important. Given its name, it should not come as a surprise that a DoS attack is aimed primarily at ensuring that the service, a computing infrastructure usually delivers, is negatively affected in some way. This type of attack does not involve breaking into the target system. Usually a successful DoS attack reduces the quality of the service delivered by some measurable degree, often to the point where the target infrastructure of the DoS attack cannot deliver a service at all. A common perception is that the target of a DoS attack is a server, though this is not always the case. The fundamental objective of a DoS attack is to degrade service, whether it be hosted by a single server or delivered by an entire network infrastructure.

5.4.1 Laying the groundwork: DoS

Before the DDoS hue and cry rose to almost thunderous proportions, DoS attacks had been tirelessly aimed at networks for some time. DoS attacks are conducted using software written to deliberately cause degradation in the target systems service levels. A number of well-documented types and variants of DoS attacks currently are available on different dubious web-pages in the wide Internet. One of the significant problems of DoS attacks is the number of freely available programs that turn this technical exploit into a task that requires the use of a mouse, a clicking finger, and a trivial amount of grey matter. This simplification can turn an Internet neophyte into a cyber criminal. A DoS attack attempts to reduce the ability of a site to service clients, be they physical users or other computer systems. This can be achieved by either overloading the ability of the target network or server to handle incoming traffic or by sending network packets that cause target systems and networks to behave unpredictably. Unfortunately for the administrator, unpredictable behaviour usually translates into a hung or crashed system [4].

Numerous forms of DoS attacks exist, some of which can be difficult to detect or deflect. Within weeks or months of the appearance of a new attack, subtle copycat variations along the same theme begin appearing elsewhere. By this stage, not only must defenses be developed for the primary attack, but also for its more distant cousins. Many DoS attacks take place across a network, with the perpetrator seeking to take advantage of the lack of integrated security within the current iteration of Internet Protocol, IP version 4 (IPv4). Hackers are fully aware that security considerations have been passed on to higher-level protocols and applications. An attempt to rectify this problem has resulted in IP version 6 (IPv6), which includes a means of validating the source of packets and their integrity by using an authentication header. Although the continuing improvement

of IP is critical, it does not resolve today's problems because IPv6 is not in widespread use.

DoS attacks do not only originate from remote systems, but also locally to the machine. Local DoS attacks are generally easier to locate and rectify because the parameters of the problem space are well defined (local to the host). A common example of a local based DoS attack includes fork bombs that repeatedly spawn processes to consume system resources. Although DoS attacks do not in themselves generate a risk to confidential or sensitive data, they can act as an effective tool to mask other more intrusive activities that could take place simultaneously. Although administrators and security officers are attempting to rectify what they perceive to be the main problem, the real penetration could be happening elsewhere. In the confusion and chaos that accompanies system crashes and integrity breaches, experienced hackers can slip in undetected.

The financial and publicity implications of an effective DoS attack are hard to measure. At best, they are embarrassing and at worst, a death blow. In the world of e-commerce, a customer's loyalty is fleeting. If a site is inaccessible or unresponsive, an alternate virtual shop front is only a few clicks away. Companies reliant on Internet traffic and e-purchases are at particular risk from DoS and DDoS attacks. The Web site is the engine that drives e-commerce, and customers are won or lost on the basis of the site's availability and speed. A hacker, regardless of motive, knows that the real place to hurt an e-business is to affect its Internet presence in some way. Unfortunately, DoS attacks can be an efficient means of achieving this.

5.4.2 Resource consumption attacks

Computing resources are by their very nature finite. Administrators around the world bemoan the fact that their infrastructure lacks network bandwidth, CPU cycles, RAM, and secondary storage. Invariably the lack of these resources leads to some form of service degradation the computing infrastructure delivers to the clients. The reality of having finite resources is highlighted even further when an attack is started to consume these precious resources. The consumption of resources (and in this instance bandwidth is considered to be a resource) involves the reduction of available resources, whatever their nature, by using a directed attack. One of the more common forms of DoS attack targets network bandwidth. In particular, Internet connections and the supporting devices are prime targets of this type of attack due to their limited bandwidth and visibility to the rest of the Internet community. Very few businesses are in the fortunate position where they have too much Internet bandwidth. When a business relies on the ability to service client requests quickly and efficiently, a bandwidth consumption attack can show, how effectively that bandwidth can be, used to bring the company to its knees [4].

Resource consumption attacks predominantly originate from outside the local network, but do not rule out the possibility that the attack is from within. These attacks usually take the form of a large number of packets directed at the victim, a technique commonly known as flooding. A target network can also be flooded when an attacker has more available bandwidth then the victim and overwhelms the victim with pure brute force. This situation is less likely to happen on a one-to-one basis if the target is a medium-sized

e-commerce site because they will, in most cases, have a larger "pipe" than their attackers. On the other hand, the availability of broadband connectivity has driven highspeed Internet access into the homes of users around the world. This has increased the likelihood of this type of attack as home users replace their analogue modems for ADSL and cable modem technologies.

Another way of consuming bandwidth is to enlist the aid of loosely configured networks, causing them to send traffic directed at the victim. If enough networks can be duped into this type of behaviour, the victim's network can be flooded with relative ease. These types of attacks are often called amplification attacks.

Other forms of resource consumption can include the reduction of connections available to legitimate users and the reduction of system resources available to the host operating system itself. Denial of service is a very broad term, and consequently some exploits cross the boundary into DoS attacks due to the circumstances surrounding their manifestation. A classic example of this scenario was the Melissa virus, which proliferated so swiftly that it consumed network resources resulting in a Denial of Service in some cases.

Anatomy of a SYN flood attack

In September 1996, a DoS attack caused a New York ISP to be unavailable for almost a week. The impact of the outage affected close to 6,000 users and 1,000 companies. The attack leveraged a technical vulnerability in the TCP/IP-protocol, that had been known for some time and was one of the first high-profile attacks to exploit SYN flooding.

A SYN flood attack achieves its desired impact by manipulating the mechanics of how a TCP connection is initiated. Unlike the User Datagram Protocol (UDP), communication streams established with the TCP protocol are connection-oriented. This means that a session must be established between the source and target computers before data can be exchanged between them. Establishing the session involves a three-way handshake, with each step commencing only when the previous one is complete.

The steps involved in the TCP three-way handshake between two machines (the client and server) can be described as follows:

1. A SYN is sent from the client machine to the server: A SYN (synchronize) packet is sent from a port on the client machine to a specific port on the server that is waiting for client connections. An Initial Sequence Number (ISN) is also submitted with the packet. TCP is a reliable protocol and consequently needs a mechanism for recovering from transmission failures and to help with packet reassembly. The ISN helps the recipient to sequence packets correctly.

2. A SYN/ACK is sent from the server to the client: The server responds to the client by sending back the client's ISN plus 1.The server's ACK acknowledges the clients SYN and the server's SYN indicates to the client that the server is able to establish a session with the client. The SYN sent from the server to the client contains the server's own ISN, which is different than the client's ISN.

3. An ACK is sent from the client back to the server: The client responds to the server's SYN/ACK with an ACK containing the server's ISN plus 1. The client and server have now established a TCP connection.

A SYN flood attack works by starting the TCP handshake by sending a SYN to the target server. The most important difference between this SYN and one originating from a legitimate user is that the source address has been spoofed. A spoofed address is an address that has been changed from the original address to another address, usually for malicious or covert purposes. The nature of IPv4 ensures that after a spoofed packet has left the source host and begins to be routed; tracing it back is very difficult, making it a favourite technique employed by hackers.

Accordingly this means now that the SYN sent from the hacker's machine during Step 1 of the handshake does not contain his real address as the source of the SYN. The address used in forging the SYN is usually a non existent address or a non routable address. IP addresses not routable over the Internet include the private IP addresses in the Class A range from 10.0.0.1 to 10.255.255.254, in the Class B range from 172.16.0.1 to 172.31.255.254, and the Class C range from 192.168.0.1 to 192.168.255.254.

The server receiving the spoofed SYN then attempts to respond to the nonexistent address with a SYN/ACK. Due to the (sometimes unreliable) nature of network connections, many implementations of TCP/IP protocol stacks are configured to wait a certain period before assuming that the SYN/ACK will not receive a response. Because the source address included in the initial SYN was forged with a nonexistent address, the server will never receive an ACK in response. In other words, the third step from the handshaking process never happens in a SYN flood attack. The connection is then left in what can be termed a half-open state [4].

A connection queue is responsible for managing the attempted connections on the server, allowing only a certain number of half-open connections to build up before future attempts to connect to that port are discarded. Only a limited amount of resources are assigned to the number of SYN/ACKs that can be queued at any one time, and the connection queue is quickly exhausted and legitimate users can no longer establish a TCP connection. A successful SYN flood attack ensures that more spoofed SYNs are sent to the server than can be released from the connection queue, effectively causing the connection queue to overflow.

A SYN flood usually involves a number of packets being directed at the target server, consequently overloading the connection buffer. Unfortunately the SYN flood attack can be quite effective, primarily because it can be launched by a hacker with limited resources and has the added advantage of obscuring the source of the attack in the first place.

Other clever twists to the SYN flood attack can include spoofing the source of the SYN in Step 1 with a legitimate routable address. Administrators observing this behaviour could then be forced to filter traffic emanating from the spoofed address, even though they are in fact not the originator of the attack. That could mean that an administrator may be faced with the task of filtering traffic coming from a branch office, partner, or legitimate user.

Anatomy of an amplification attack

An amplification attack achieves its effectiveness by enlisting the aid of other networks that act as amplifiers for the attack. This allows hackers with limited resources to target victims with a considerable increase in resources. The networks used in the amplification attacks are usually oblivious to their part in the whole process.

Two examples of amplification attacks are the strangely named Smurf and Fraggle-attacks. Unfortunately, the only harmless elements to these attacks are their names. The Smurf attack gained its name from a program that leverages this particular attack methodology. A Smurf attack is staged by using a combination of loosely configured networks and the Internet Control Message Protocol (ICMP). IP was not designed to be reliable and consequently requires a method of providing status and error information. This is where ICMP steps in. ICMP is used for, amongst other things, error control. The well known ping command uses ICMP to determine if a host is alive by sending an ICMP echo request to a host. If the host is up and running a TCP/IP stack, it replies with an ICMP echo reply. A Smurf attack exploits this seemingly simple dialogue by spoofing the source address of the initial ICMP echo request [4].

The first step in the process is for the attacker to place the victim's IP address in the source address field of the ICMP echo requests. The destination of the ICMP echo request can then be any "loosely" configured network that has a router that broadcasts to its subnet, and similarly, hosts that will respond to the echoes on the network broadcast address after they have passed through the router. This may in itself sound relatively harmless, but a couple of factors exacerbate the problem. First, the attacker sends the ICMP echo not to a specific IP host, but to the broadcast address of the loosely configured network. Sending an ICMP echo request to a broadcast address of a network causes the echo to be processed by every machine on that network.

Considering a scenario in which fifty hosts are assigned network addresses within the IP range 192.168.100.1 through to 192.168.100.254 and a subnet mask of 255.255.255.0, all machines on this network will respond with an ICMP echo reply, if the following simple command is issued:

```
ping 192.168.100.255
```

The single ping command then causes 50 responses directed at the client considered to have issued the command. In other words, the original message has been amplified 50-fold! How does this form of amplification relate to the Smurf attack? The machines on the loosely configured network will then respond to ICMP echoes with an ICMP echo reply directed at the spoofed address. In other words, the victim becomes the recipient of the replies to the ICMP echo. Secondly, the attacker usually ensures that he sends a number of ICMP echoes. The victim then receives ICMP echo replies equivalent to the number of original ICMP echoes sent by the hacker, multiplied by the number of hosts on the broadcast address. If two hundred hosts are on the broadcast address, then the attacker could magnify a single ICMP echo into 200 ICMP echo replies (see figure 5.2).

In this example we have simplified the context of the attack by assuming that the hacker has used a single loosely configured network to act as an amplifier. If an attacker uses

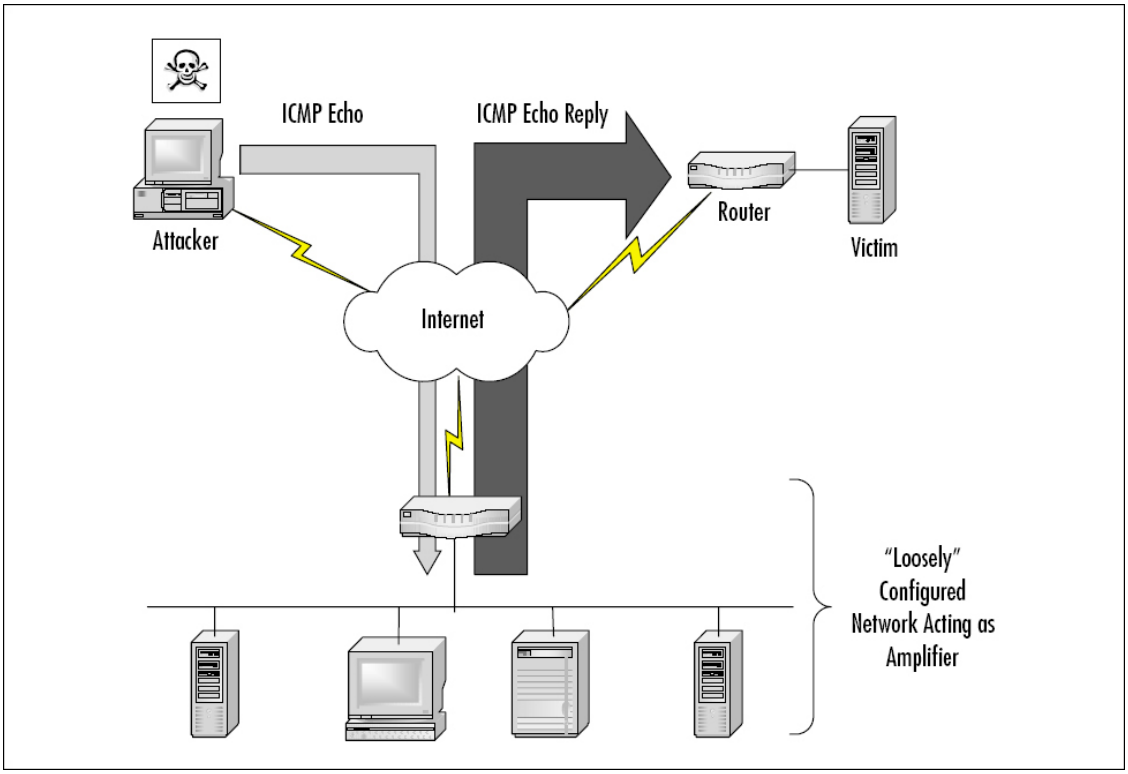

Figure 5.2: Composition of a smurf attack (Source: Hack proofing your e-Commerce Site)

multiple networks, the traffic generated would be larger and more diverse and therefore harder to filter.

The Fraggle attack is a variant to the Smurf, exploiting similar amplification methods by directing UDP packets to network broadcast addresses. Fraggle relies on the largely unused UDP services chargen and echo. The amplification network used by the Fraggle attack responds to the UDP packets by sending UDP messages to the spoofed address.

A side effect of amplification attacks is that they can affect two victims: the amplifier and the owner of the spoofed address. The network the attacker used to bounce the ICMP echo experiences similar problems as the final victim, such as network congestion, slow response, and possibly a total denial of service.

5.4.3 Malformed packet attacks

Operating Systems (OSs) have a notorious reputation for falling over at the slightest provocation. Considering the variety of uses the modern OS is put to, they perform quite well. Even though they are pushed through rigorous testing cycles and patched on a regular basis, they can behave unexpectedly when non standard events occur.

For the hacker interested in DoS attacks, an unexpected situation hopefully leads to resource contention or a crashed system. A malformed packet attack usually consists of a small number of packets directed at a target server or device. The packets are constructed

in such a fashion that on receipt of the packet, the target panics. A panic is considered to occur when the device or operating system enters an unstable state potentially resulting in a system crash.

A classic DoS malformed packet attack is the **Ping of Death**. Most vendors of network hardware and software have been hardened to what was once the scourge of the Internet community. The Ping of Death consists of directing a large ICMP echo at the victim. The ICMP echo can be generated using the ping command, but the packet size must exceed 65535 bytes, which is the maximum size of an IP packet, or contain 65507 bytes of data. The ICMP packet is not transmitted "as is" and may be broken up because the underlying transport has a smaller maximum packet size. For example, the maximum packet size for Ethernet is typically 1500 bytes. On reassembly at the target, the ICMP echo overflows the OS buffer (which is not expecting a packet larger than 65535 bytes), causing the machine to crash or become unstable [4].

A typical Ping of Death command could look like this:

```
ping -1 65515 victims.address.com
```

A number of variations along similar lines to the Ping of Death are in circulation, many of which vendors have supplied fixes for. Included in this list are:

- **Teardrop**: This attack exploits a vulnerability during the reassembly of IP packets on target hosts. Large packets are fragmented into smaller packets that need to be reassembled at the target. The fragments include an offset to the beginning of the first packet that enables the entire packet to be reassembled. In the Teardrop attack, the offsets are changed, making it impossible for the target system to reassemble the packet properly. This unexpected situation causes the OS to become unstable.
- **Bonk/Boink**: This attack exploits the reassembly of malformed UDP datagrams.
- **Land**: This attack sends a malformed packet during the setup of the three-way TCP handshake. The initial SYN is sent to the target with the victim's address detailed as both source and destination.
- **Malformed RPC**: This attack utilizes malformed RPC packets to disable RPC services.

5.4.4 Anatomy of a DDoS attack

Though some forms of DoS attacks can be amplified by multiple intermediaries, the first step of a DoS exploit still originates from a single machine. DDoS attacks advance the DoS variant one more painful step forward. DoS attacks have evolved beyond single-tier (SYN flood) and two-tier (Smurf) attacks. Modern attack methodologies have now embraced the world of distributed multi-tier computing. One of the significant differences in methodology of a DDoS attack is that it consists of two distinct phases. During the first phase, the perpetrator compromises computers scattered across the Internet and

installs specialized software on these hosts to aid in the attack. In the second phase, the compromised hosts, referred to as zombies, are then instructed through intermediaries (called masters) to commence the attack (see figure 5.3).

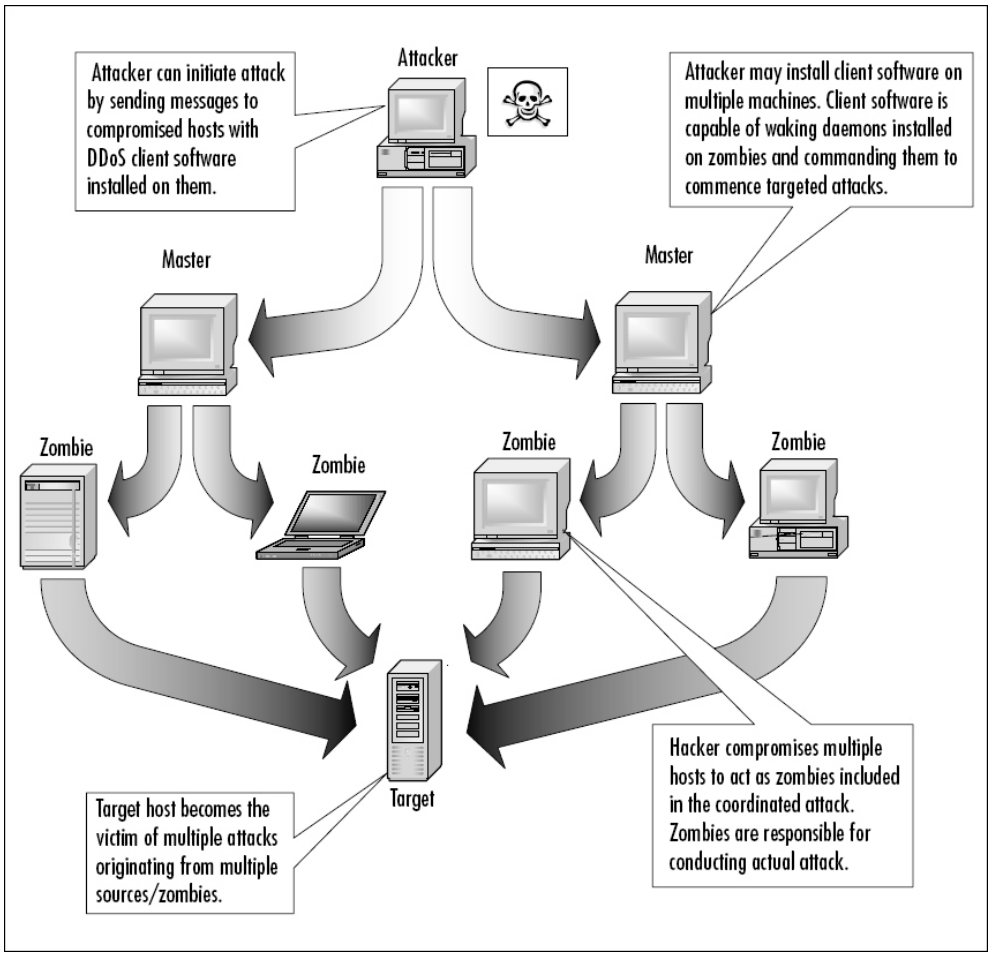

Figure 5.3: Composition of a denial of service attack (Source: Hack proofing your e-Commerce Site)

Hundreds, possibly thousands, of zombies can be co-opted into the attack by diligent hackers. Using the control software, each of these zombies can then be used to mount its own DoS attack on the target. The cumulative effect of the zombie attack is to overwhelm the victim with either massive amounts of traffic or to exhaust resources such as connection queues.

Additionally, this type of attack obfuscates the source of the original attacker: the commander of the zombie hordes. The multi-tier model of DDoS attacks and their ability to spoof packets and to encrypt communications can make tracking down the real offender a tortuous process. The command structure supporting a DDoS attack can be quite convoluted, and it can be difficult to determine a terminology that describes it clearly. Perhaps one of the more understandable naming conventions for a DDoS-attack structure and the components involved is detailed below.

Software components involved in a DDoS attack include:

- **Client**: The control software used by the hacker to launch attacks. The client directs command strings to its subordinate hosts.
- **Daemon**: Software programs running on a zombie that receives incoming client command strings and acts on them accordingly. The daemon is the process responsible for actually implementing the attack detailed in the command strings.

Hosts involved in a DDoS attack include:

- **Master**: A computer from which the client software is run.
- **Zombie**: A subordinate host that runs the daemon process.
- **Target**: The recipient of the attack.

In order to recruit hosts for the attack, hackers target inadequately secured machines connected in some form to the Internet. Hackers use various inspection techniques, both automated and manual, to uncover inadequately secured networks and hosts. Automated trawling for insecure hosts is usually scripted and can, under the correct circumstances, be detected by a company's security infrastructure. Depending on the hackers' level of competence, manual inspection can be harder to identify because the attacker can adapt his approach accordingly, but it is also much more time consuming.

After the insecure machines have been identified, the attacker compromises the systems. Hackers gain access (root, usually) to a host in a startling variety of ways, most of which, quite sadly, are preventable. The first task a thorough hacker undertakes is to erase evidence that the system has been compromised and also to ensure that the compromised host would pass a cursory examination. The tools used to ensure that these tasks will be successful are sometimes collectively called rootkits [4].

Some of the compromised hosts become masters while others are destined for zombification. Masters are installed with a copy of the client software and are used as intermediaries between the attacker and the zombies. Masters receive orders that they then trickle through to the zombies for which they are responsible. Available network bandwidth is not a priority for hosts designated to be masters. The master is only responsible for sending and receiving short control messages, making lower bandwidth networks just as suitable as higher bandwidth networks.

On the hosts not designated as masters, the hacker installs the software (called a daemon) used to send out attack streams and the host graduates to become a zombie. The daemon runs in the background on the zombie, waiting for a message to activate the exploit software and launch an attack targeted at the designated victim. A daemon may be able to launch multiple types of attacks, such as UDP or SYN floods. Combined with the ability to use spoofing, the daemon can prove to be a very flexible and powerful attack tool.

After the attacker has recruited what he thinks are a sufficient number of zombies and has identified his victim, the attacker can contact the masters (either via his own methods

or with a specially written program supplied with the DDoS program) and instruct them to launch a particular attack. The master then passes on these instructions to multiple zombies who commence the DDoS attack. After the attack network is in place, it can take only a few moments to launch a distributed attack. With similar speed, the hacker can also halt the attack. The basic flow of the attack then becomes:

For hosts: Attacker → Master → Zombie → Target

For software: Attacker → Client → Daemon → Target

To provide a context for the possible scale of DDoS attacks: A DDoS attack was mounted on the University of Minnesota by hundreds of zombies that denied afterwards network access to thousands of users for three days. In fact, also Microsoft became a victim in the line of bemused businesses subjected to successful DDoS attacks. The use and development of DDoS programs have won the interest of governments, businesses, and security experts alike, in no small part because it is a new class of attack that is extremely effective while simultaneously being hard to trace [4].

5.5 Actual threat

Every year the CSO-Magazine makes surveys about relevant internet crime. 70 percent of the questioned businesses recorded an attack in the year 2003. The number of enterprises with an increasing attack rate laid with good 40 percent higher than in the previous year, only 6 percent reported a reduction of attacks [9]. Over 80 percent head losses: more than the half reported operative loss, a quarter marked financial detriment and about 10 percent noted other losses [9].

A damage of 666 million dollars was created by the 500 questioned companies, as a consequence of electronic crimes in the year 2003. 50 percent of the answering businesses do not know how high the damage is.

These statements point at a significant increase from attacks and require measures in order to dam the economic damages. In barely the half of the electronic attacks was the law called in to help [9]. A cause could be that the businesses fear to damage their reputation if they report e-crimes. They favour many times collaborator supervision systems despite possible injury of the data protection. Surely another cause could be that the IT-responsibles do not know the accurate law situation in the concerned cases [9] and because of that they can't respond adequately. In many times the law situation is unclear. Relating to this, on the one hand a global base demanded, on the other hand it is necessary to create more clarity.

5.5.1 Types of electronic crimes

The eCrime Watch of Summary offers a listing of feared attacks on questioned businesses (see figure 5.4): 77 percent of the businesses reported virus attacks. Interesting is on the

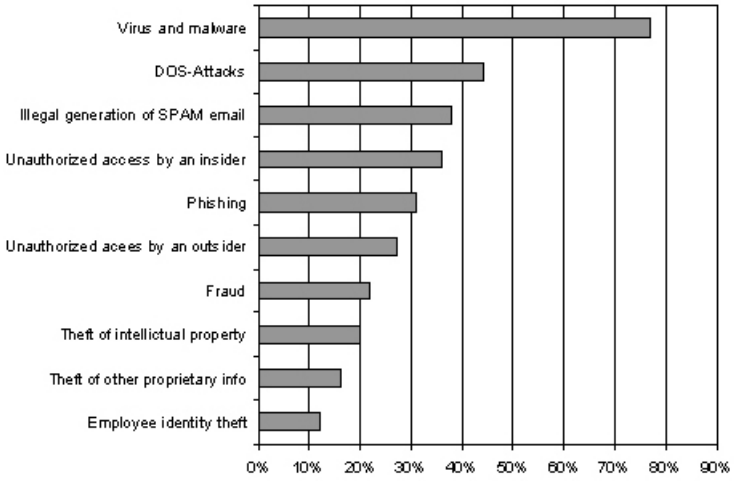

Figure 5.4: Electronic Crimes (Source: eCrime watch summary 2004)

second place are Denial of service attacks with 44 percent of the questioned businesses. In the following we will deal with DDoS-attacks above all because they are threat do not only depend on the configuration of the threatened server but also from the security of other computers in the internet; that can be abused by malware for the installation. Consequently, the measures against DDoS-attacks grasp also for viruses and Trojan horses.

The type of attacks can be subdivided after impacts as follows (see figure 5.5):

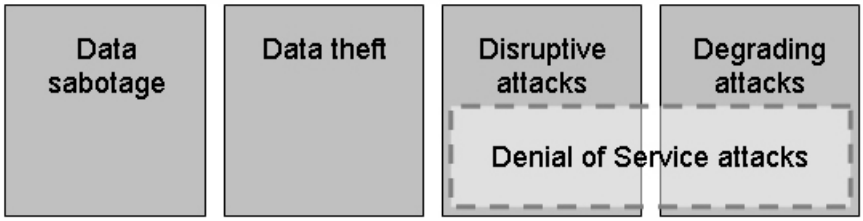

Figure 5.5: Impact of electronic crimes (Source: Lukas Keller)

At the data destruction, data are deleted. If there are no backup copies available, this will emerge an immense damage. If data are stolen, a loss can be created that the adversary takes personal advantage from the stolen information or uses it against the company.

Since 1999 the increasing integration of the internet causes with distributed denied of service attacks (DDoS-attacks) a new impact. The goal of DOS-attacks is to slow down the server or to steal its resources; they do not aim at stealing data.

There are two types of DDOS attacks [7]:

The attacker tries in so-called **disruptive Attacks** to decrease the availability of the server so that that does not react on questions of the Clients. Servers are speeded down by DDoS-Attacks to blackmail protection money of the content provider for example. Experts of the British NHT-CU suspect systematic action of the Russian or East European Mafia since the heights of the protection moneys and the action correspond to the custom of professional blackmailers in this area [11].

The aggressor tries in **degrading attacks** to consume a part of the resources of a server. These attacks actually don't lead to a total server cancellation and therefore remain many times unnoticed, at least for a certain period. However such attacks can cause big losses since the server resources are laid out on periods of maximum load; and through the degrading attack the server can't withstand. If the victim doesn't recognise the attack, he updates the resources instead to remove the varmint.

Through DDOS-attacks, big costs can originate if the servers are overloaded. The company can be restricted in its operative activity, web Portales cannot put for the supply-chain and for the customers to disposal if necessary.

Consequently, an attack possibly damages the customer relationships, the business reputation and decreases the effectiveness of the organization to the suppliers. At least theoretically a DDoS attack can threaten a business seriously.

An example: in 2003 the airline of Easy Jet almost generates a turnover of barely 1 billion £ exclusively with the sale of tickets. Easy Jet expels 95 percent from it online [13]. In consequence of a server breakdown in case of an attack the loss would be 250'000 £ a day. With the expelled profit of 52 million £ Easy Jet would write losses within 20 days.

5.5.2 Attackers identity

After evaluations of the eCrime Watch of Summary , 40 percent of the interviewees put hacker attacks as the biggest threat. 22 percent mention current employees and 6 percent called former employees as the biggest electronic threat. As further impact are considered the competition, customers, information brokers, service providers and finally terrorists (see figure 5.6).

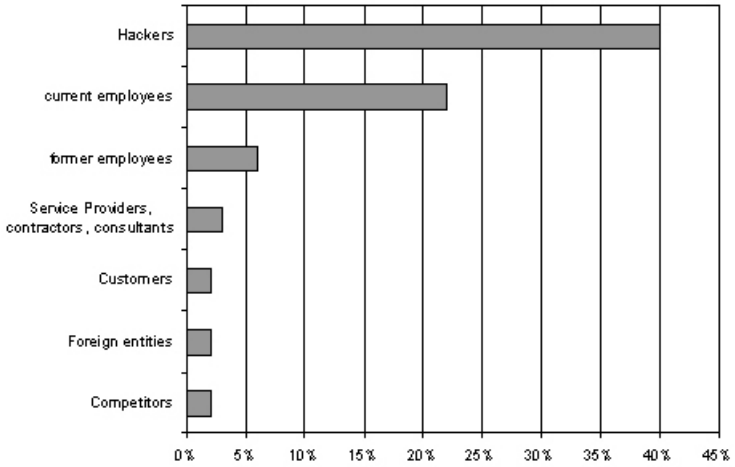

Figure 5.6: Cyber security threat (Source: eCrime watch summary 2004)

According to experience a third of the attacks comes form insiders, and 60 percent of the attacks from outsiders [9].

The identity of the aggressors of DDoS-attacks is difficult to investigate. On the one hand the sender addresses (IP-Spoofing) are falsified, on the other hand the attacker operate indirectly over miscellaneous, insufficiently protected computers, wherefrom they attack the victim. There are three main groups of attackers amounted, which can be classified accorded to motivation and professionalism (see figure 5.7).

On the one hand, script kiddies drive their nuisance, which search the net more or less focussed with scripts for vulnerable victims and starts attacks.

There are tightly focussed DDoS attacks for example against Computerbetrug.de, where a hacker was offended by the webmaster. In consequence he started a DDoS-attack to demonstrate his power [14]. Another example is the entrepreneur Daniel Rödding. He estimates a DDoS-attack as the only possibility to stop a spammer [15].

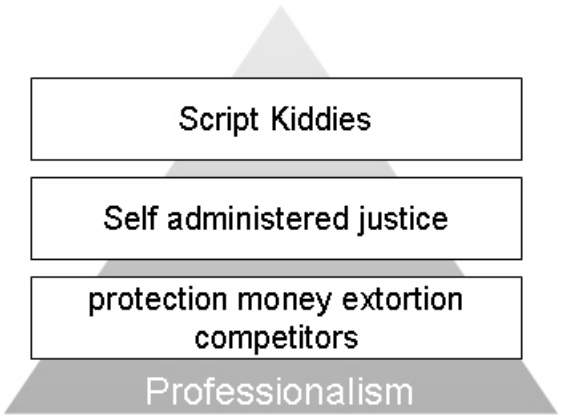

Figure 5.7: Attackers identity (Source: Lukas Keller)

DDoS-attacks on professional base are blackmailing protection money. Other possibilities are attacks from the competition to get an advantage for their self.

5.6 DDoS defense mechanisms

First, an aggressor of a DDoS-attack procures the access to miscellaneous computers in the internet. There, he installs programs, which flood a server on command with inquiries. The concept is essentially based on two weak places: namely IP-spoofing (falsified sender addresses of the packets) and many insufficiently protected internet computers, on which forbidden-proves, the attack programs were installed.

The special thing about these DDoS-attacks is that also those, who are protected optimally from internet invaders, can be hit. In this respect, insufficiently protected computers are a danger not only for the respective operator but also for all other computers in the internet.

In the following section we respond on possible activities for the protection against DDoS-attacks. We relate to the taxonomy of ucla-report [7]. We divide the countermeasures into preventive and reactive. The preventive ones mark the conducted security mechanisms

independent of an attack, the reactive countermeasures are conducted in a case of an attack.

5.6.1 Preventive countermeasures

Attack prevention mechanisms are modifications at the system configuration; that avert DDoS-attacks. Attack prevention mechanisms are subdivided in system security in and protocol security mechanisms (see figure 5.8).

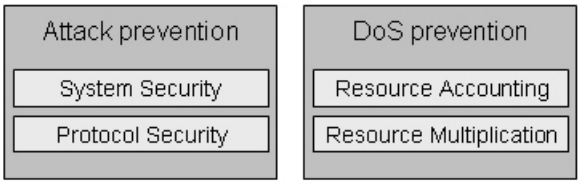

Figure 5.8: Preventive DDoS defense mechanisms (Source: Ucla tech report)

System security

System security mechanisms include the complete protection of the system like access protection, removing application bugs etc.

Examples for system security mechanisms are security patches, firewall systems, virus scanners, etc. such systems decrease the strength of DDoS-attacks, but can't remove the problem completely.

Protocol security

Bad protocols, that are resource-costly on the part of servers and cause many transactions, are another weak place. A possibility for aggressors is to incomplete packages to the server. The server now tries to complete the packages. Also the TCP SYN attack is known: the aggressor bombards the victim with SYN-packages and doesn't answer with an ACK to confirm the connection. With badly implemented TCP-protocols, the server wastes his resources on trying to build up connections for all the SYN-packages.

DDoS prevention mechanisms aim on serving further clients during an attack but not to get overloaded.

Resource accounting

Resource accounting mechanisms part the resources amongst the users depending on their privileges. Unauthorised users have no access to the resources.

Resource multiplication

Another strategy is to distribute the resources on miscellaneous servers and to multiply the resource consequently. So, a DDoS-attack must be bigger to be efficient. A disadvantage of this mechanism is that it is expensive to get additional resources and to get them connected.

5.6.2 Reactive countermeasures

Also ingenious firewalls and authentification systems don't lead to a hundred percent security against attacks. Particularly, they cannot protect from attacks from the own network. It is more important, to recognize attack attempts and successful attacks, to be able to introduce countermeasures for the future, and to guarantee, that co-workers don't access fake data without knowing it (see figure 5.9).

Despite of meaningful precautions on own IT-infrastructure break-ins are possible. Break-ins should be discovered as fast as possible to restrict the damage as narrowly as possible. In the ideal case, the attack should be discovered and counter-measures introduced before damage can be created.

Figure 5.9: Reactive DDoS defense mechanisms (Source: Ucla tech report)

Intrusion detection systems (IDS) offer us following procedures [12]:

Pattern attack recognitions systems

Pattern attack recognitions systems are based on the fact that an attack leaves typical tracks in each case. The pattern of an attack gets stored in a database and gets compared with the continuously track in the system. But the system recognizes only attacks with similar patterns like the ones in the database. Many of these signatures can only be assessed as indications of an attack, if they are cumulated or appear in uncommon context. For example a try to login with a wrong password is certain no attack, but a couple hundred tries attempt to be one. One can distinguish the pure pattern attack model of an threshold controlled one.

Advantages of the signature recognition are:

- The expenditure for the installation and maintenance is low, if signatures stand by disposal
- The Attacks are frequent executed by attack scripts, from which signatures can be derived easily, so that high discovery probability is given.

Disadvantages are:

- New attacks are not recognized. The signature database must be adapted on a regular base to new discovered attacks. The success depends directly on the quality of the signature database
- An adaptation of the signature data base to local realities or a definition of new patterns is usually very costly. Therefore, such data bases are limited portable.

Anomaly attack detection systems

Anomaly attack detection systems measure deviations of the expected use of a resource. The expected use is stored as standard value. If the deviation is big enough, the system will report it. In a refined base, also the expected sequence of events can be considered: If the sequence of the events runs out differently as expected, the system assesses this as attack.

Anomaly attack detection systems have following advantages:

- So far unknown attacks can be discovered
- It can recognize attackers, who work under wrong user account but without causing explicit abuse.

Disadvantages are:

- The election of the parameter attitude is critically and can easily lead to false alarms or not recognize attacks.
- Anomaly recognition still is object of the current research and far from the practical employable ness distant.

It makes sense to start hybrid IDS, with anomaly and pattern recognition, to use the advantages of both systems.

Intrusion response system (IRS)

Intrusion response systems aim on minimizing the damage with the victim, by keeping the connection to the clients upright and not simply separate the connection. The ucla tec report distinguishes the following response-strategies:

Agent identification mechanisms

Agent identification mechanisms identify the aggressor: If the attack seems to pursue a concrete goal, one should try to identify the aggressor [12]. This necessitates a high expenditure that must be not absolutely justifiable. The selection of the correct time for a widened logging can become well certain through an IRS. To be able to guarantee the security and the confidentiality of the own data during the identification of the aggressor, it is possible to lure this into a "Playground", that can be put by the IRS to disposal. This will lead the aggressor to a special computer, which seemingly contains valuable, but there he can't cause any damage. If the aggressor is identified, its obvious to start a counterattack which miscellaneous IRS programs intend. However such a counterattack is problematic about following reasons:

- The counterattack must be suitably and legally justified
- The identified computer should agree with the computer of the aggressor, a problem for DDoS-attacks is that zombies and not the aggressor themselves attack the victim.

Rate limiting systems

Rate-Limiting of system limit the spread of the streams, that is classified as malicious from the IDS. The advantage of this procedure is that the attack-streams get limited not and the data flow doesn't get completely switched off, if the IDS shows a high false recognition rate. The disadvantage is that in case of an attack, it lets pass the attack-stream, even if limited. So, an attack can nevertheless be effective and restrict the accessibility of the clients.

Filtering systems

Filtering-Systems are suitable, if the IDS can identify the attack-stream exactly. After that the recognised packages get filtered out. In the principle, this corresponds to a dynamic Firewall. A disadvantage of this mechanism is, if the IDS classifies allowed packages as malicious and filters them out in consequence.

Reconfiguration mechanisms

Reconfiguration Mechanisms aim at altering the topology of the victim or the immediate surroundings. In the case one attack more resources can be assigned to the victim, to manage additional inquiries, or the victim gets isolated. Reactive DDoS defense strategies work differently, autonomously or cooperative:

Autonomous mechanisms

Autonomous mechanisms work independently from others. As for example Firewalls or IDS do.

Cooperative mechanisms

Cooperative mechanisms work independently of each other, but the result could be better, if they cooperate with others. Dynamic Firewalls, that can block additional malicious data streams with piece of information from the IDS, are an example of it.

Interdependent mechanisms

Interdependent mechanisms do not operate autonomously. An example for that are trace back mechanisms to identify the attacker. But they make no sense, if they are installed only on one single router.

5.7 Suggested countermeasures and costs

Generally, the danger of attacks cannot be prevented by 100 percent with preventive measures. There remains a risk always, which causes costs. Especially DDoS-attacks contain this problem clearly. Also optimally from invaders protected systems can be effected by DDoS attacks.

	Preventive Measures	No preventive Measures
Preventive Costs	Yes	None
Reactive Costs	Manageable	Non-manageable

Figure 5.10: Reactive costs (Source: Lukas Keller)

This remaining risk is to minimize to hold the emerging damage as small as possible. In the chapter "preventive countermeasures" are further implementations located which measures should be taken. If no measures are taken, the server remains unprotected to the attack and the consequences are not manageable (see figure 5.10). Questions about the possible insurance of damages through electronic crime are certainly interesting.

Effective steps of distributed Denial-of-Service-attacks must be taken consequently at many places with miscellaneous target groups. Server operators in the internet are able to take a series of meaningful measures, but they can't solve the DOS problem completely. Rather, different target groups must become active in their area. Regarding the threat through DDoS-attacks, the internet only becomes secure if the target groups work

together. In response of this, the transaction of Denial-of-Service-attacks will get more difficult as well as a later prosecution of the originators of these attacks is eased.

For the target groups below the following proportionate actions will show up [16].

5.7.1 Target groups

End-users: Operators of computers, with which information is called in the internet.

Internet service providers: Operators of the network infrastructure (for example node computers or routers)

Server operators: Companies, that administer and configure servers. The servers offer services and information in the internet (WWW-Server, DNS-Server, MTAs, Proxy-Server...)

Content-provider Producers of editorial contents that are prepared by server operators in the internet for example

5.7.2 Countermeasures for the target groups

Figure 5.11 shows an overview of the Countermeasures for the target groups. It is applicable to all target groups that the employees are educated sufficiently. So, risks can be prevented through prevention on the one hand, and on the other in a critical case they act adequate to minimize the damage. Responsible employees and administrators must be in their specific field up to date so that they can use the disposed Tools optimally. According to a study by Ernst & Young only 50 percent of the examined companies put their employees on a security awareness training. If the people is considered as security-critical point, then, the network becomes weaker the more people are interconnected. The most effective means are also after Ernst & Young investments in a security culture [17].

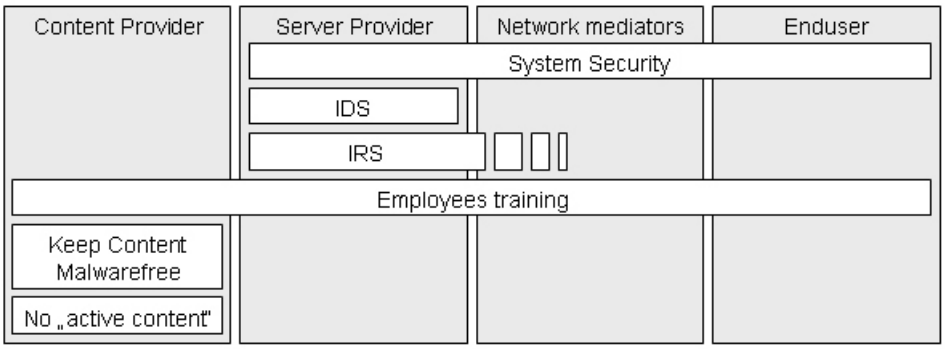

Figure 5.11: Measures for the target groups (Source: Lukas Keller)

Measures for content-provider

The content provider should select primarily server operators after security and availability criterions. They should choose server operators who are experienced with the required internet platforms and who can prove his efforts in the area of IT-security through security concepts etc. Here arises costs through a careful evaluation and at most through higher outsourcing costs, but they get compensated through smaller cancellation risk of the server.

Through avoiding active contents and others security critical technical settings, content provider can contribute that no uncertain settings must exist on the clients. So, they impede that programs get installed unnoticed on the computers of the final users. For the content provider, no additional costs occur. Possibly, he has to give up exceptional features.

The content provider should check with particular search programs daily whether malicious programs are available on his sites. Download-sections are attractive to attackers to install Trojan horses, viruses or other malicious programs because they reach a wide public on that way. Such a procedure is tempting especially for DDoS-attackers, since a large number of computers is required for an effective attack. This measure doesn't serve primarily the content supplier but comes towards all. The content provider profits indirectly by offering a malwarefree homepage. If an attacker succeeds with installing malware on his home page, it damages the reputation of the content supplier as the case may be.

Measures for server operators

Not only are the computers of the server operators considered as victims of the DOS attacks. Because of their efficient binding to the internet, they are also potential exit platforms. Therefore it must be prevented that these computers are abused as starting point for attacks on further computers.

In the area of system security mechanisms considered by the above mentioned taxonomy filters, access protection, IDS and installation of security patches are recommended as measures for server operators.

On the router which is connected upstream to the server, should be installed a filter, that makes IP-Spoofing impossible. This filter should be configured in a way that it lets only to protocols, for which services are installed on the server, pass. It is recommended to install a dynamic filter, which filters IP-packages in the case of an attack with the addresses reported by the IDS. The industry offers Routers with filter and integrated IDS in all performance classes, as for example the Attack Mitigator of toplayer [18].

The servers of the server operators can be abused also like the computers of the end-user as an agent for DDoS attacks. For that the attacker installs malware at known weak points. Therefore the servers should be configured carefully and securely and security gaps should be patched regularly. It is at a restrictive give-away of access rights to the

users to access the prepared system resources and to watch carefully at configuration changes. At a frequent base the file system should checked for integrity. If merely static data are required, a manipulation-sure, writing-protected data carrier can be used.

If weak places are discovered in the software, it is important that these can be remedied as fast as possible. If the source code of software is freely available, frequently the weak places are remedied faster than in other products. Therefore, Open-Source-products with similar cost-quality-relationship should be preferred, whereas open source products cause mainly keep and care costs.

The system should be guarded by a intrusion detection system. In the case of an attack the IDS will detect access attempts with a big likelihood and will report IP-addresses at a filter, which will block them (see DDoS defense mechanisms) in case of an attack it is of central importance, to react as fast and right as possible. Only that way it is possible to introduce effective counter-measures which eventually identify the attacker and to put up the normal business in short time to reduce the damage. Therefore it is to define a suitable emergency plan, which makes a structured procedure possible and a which generates a formalized reporting. Necessary statements for the emergency plan includes among other thins support, alternative communication ways, act instructions, responsible people, and storage place possibly required resources. Only 50 percent of the companies questioned by NHTCU [10] dispose a formalized method, if an electronic crime is identified. In the finance sector, the number lies by 70 percent.

Measures for Internet service providers

Generally, the Internet service providers are not a target of a DOS attacks. On the one hand they have indirectly benefits of a secure internet, since the trust of all users grows and therefore the number of users increases. On the other hand, they have directly operative benefits to keep their availability upright through suitable measures, because the network connections can be incriminated so strongly through an attack, that their bandwidth is exhausted. Effective mechanisms prevent the overload of the network consequently and spare the purchase of further resources through it.

Internet service providers can help to restrict the IP-spoofing [16]. An organization that is connected to the network operator has a certain IP-address area to disposal. Each IP-package of this organization now should come from this address area. If that is not the case, the IP-address is faked, and the package should get filtered by the network operator. So IP-Spoofing becomes strongly restricted and can only take place in the assigned address area. Over normal home accesses, IP-Spoofing would be not possible anymore, because it only gets one IP-address assigned.

The filtering could get connected with dynamically at the server operators IDS to filter current attacks as early as possible.

Measures for end-users

Computers of the end-user are not goal of DOS attacks in the normal case. However, they can be abused for attacks as an agent. Therefore, also final users make a contribution to the protection of DOS attacks, in which you perform measures for the system security (firewalls, virus protection) and switch off active contents in the browser. None direct costs occur for the user, there are free effective security tools for the private use like Firewalls and virus protection. For companies these are available for a low price. For the end-user occurs more expenditure, but the security measures will constrict the effect of viruses and malware. Consequently, the higher expenditure is compensated by system breakdowns cause of viruses.

5.8 Possible future development

Describing the future is quite a hard job. It's not possible to forecast the conditions in perhaps ten years, but there are another ways that allow you to get some kind of an idea how it could continue. Knowing a little bit about that would also be needful for all computer users like enterprises, server maintainers as well as home users, because they possibly get a better chance to prepare and prevent from some kind of attacks. But how you can get a crude picture of the future? First of all, you can look at the past and analyze, how malicious attacks and the financial consequences for enterprises developed in the last years. How it was earlier, what do we have today? With experience you have the chance to see some trends, to suggest what could be coming up in the next time! The following part is mainly based on experts meanings and statements about the future.

5.9 Law development trends and the customers interests

E-commerce has become more and more significant, and it will even get more important in the future. But not only the financial oriented world is touched by the growing Internet applications scene. Think about, for example, GPS navigation. Most vehicles like private cars might be supported by GPS systems in a couple of years. Paper maps could get outdated considering coming up possibilities offered by small mobile clients with satellite navigation. One day it will be possible to vote on the Internet. Till we have gotten so far, there will still be more or less time.

But it is no question that global network applications are getting more important every year. Economy, environment and our own life is getting the longer the more depending on computer network systems, connected to the Internet. At the beginning, and for quite a long time, the Internet was considered as a law-free space. But certainly, there must be good laws when Internet applications are beginning to integrate in your all days life. Symposion Publishing GMBH noticed a lack of consumer protection in online banking and financial applications at the beginning of 2003. There was also no really protection

for customers, they were responsible for themselves and any financial or similar damage, caused by attacks, would often be their own problem. Data holding and safety had to be more safe, enterprises in the future may be requested by law to ensure the safety for their customers, which also includes a full insurance and good cryptographies for sensitive customer data.

5.10 Possible future development of hacker techniques and goals

It is quite hard to describe the future development of digital malicious attacks. There is much room for many surprises in the fast-living world of computers. But if you are looking to the past, how worms developed in the last years and how they are working today, you might be able to suggest possible future developments. One look at the today security techniques against virus attacks and espionage could also be helpful.

We will present a small summary of experts opinions here, sourced at Nazario (2004). How do they perceive this problem, which trends do they see for the future? A regard to future trends like this is very important. New exploits for new security holes and other attacks can happen daily, even though there are very high security levels and wide-known issues.

5.10.1 Intelligent worms

A highly efficient worm has to fulfill as many as possible of the following properties:

- **Compatibility to several hardware and software platforms**: Windows, Mac OS, Linux and more. This allows a much better spreading of the worms, because they are not specialized on just one type of system. Just that is one of the biggest challenges for hackers: there is usually no simple way testing in a wide area, because the development of worms must happen hidden. Additionally there have to be (or have to be found) special security holes in the several operating systems that are able to help this worm getting in the system. But there is no big lack of security issues in more or less operating systems on the market, so this chance keeps more or less existing permanently.
- **Well-hiding**: Another hackers target is to prevent worms on infected hosts to be found and eliminated as long as possible. Because of that, it's necessary to try to hide the broken-in worms. There are several ways to do that, one simple possibility is to give the worm a normal system service name, to be overseen by human monitors. Another possibility is to implement the worm as a kernel module, bypassing system calls so that they won't appear in any system processes list any longer. All this measures are good for undisturbed spreading of the worms, the can get active on more systems. As soon as a worm gets detected, it will be killed relatively fast -

firstly on the infected host where it was detected, short time later also on the other hosts via anti virus updates over Internet.

- **Independence**: The worm should be able to act and spread as independent as possible from human users behavior.

- **Ability to learn**: The spreading worm should be able to learn, for example learning new techniques to attack. To realize that, one or more decentralized database hosts could be used, where the worms could access independently: taking information from the server as well as send new information to that database.

- **Ability to monitor attack targets**: Worms should have a mechanism that allows them to monitor potential attack targets, testing the behavior and preferences before an attack. This action should be done non-striking and well-hidden, so that it is well-covered and not detected in the usual data stream.

- **Polymorphic design**: The worm should have polymorphic preferences (hiding signatures, cryptographies). Only for executing it has to be decoded. That helps to hide it much better from monitoring software.

- **Many different possible actions**: It should be able to take as many different actions and activities as possible in the network. The most important target is not only the spreading as far as possible, it's more the ability to run special tasks after spreading, for example the DDoS attacks.

One interesting point in such experts future forecasts is the fact, that a lot of them have really gotten real a short time later. Perhaps the hackers heard from their ideas and were inspired. In this way, the experts could even have helped in some kind developing new virus techniques [19].

5.10.2 Modular, updateable worms

Todays worms often suffer from a set of limitations that disturb their spreading and working:

- **Limited possibilities**: The possibilities the worm owns are limited. Only a few attack techniques, attack behaviors, techniques to hide on the host and so on. That limits can lead quickly to the elimination of the worm.

- **A limited set of actions to attack servers**. This automatically limits the absolute number of servers that are vulnerable for the worm. With every newly attacked server there will be one server less, until there are no more vulnerable servers left for the worm. This can be a danger for the worms spreading if it gets enclosed in a small network (for example a small island of vulnerable servers in the Internet) with no bridges that lead to the world outside.

- **Critical data traffic amount**: The more worms are spreaded, the more data traffic will they generate and the higher is getting the risk of being noticed by security monitors because of that.

- **Data traffic generated by communication**: A big problem for worms can also be the network topological system many worms are using. A generally open and intensive communication between the nodes (=infected hosts) helps monitors to detect them quite fast (because of striking data traffic amounts) and even to find out their potential origin.

- **Pretty centralized information network**: Worms with database information system often use a central database source, where for example attacked hosts, potential victims and other information are collected. Such databases can, once found by security staff, be used to find out very fast and simply which hosts were attacked. Worm was killed very fast, every host administrator can be informed, and the further spreading can be directly stopped by breaking the connections between hosts and the database center.

Hackers are thinking about these limitations, how they could be eliminated. New developments could go the following ways:

- **More passive behavior**: New infection mechanisms (passive host scan instead of an active one), only watching the generated data traffic without inquiring, finding out the server operating system and software and similar. The worm collects these informations and will probably be able to start a specific single, but successful attack.

- **New network topology**: There have to be used new network topological systems which make it harder to detect them. Models like directed tree model or Gnutella network are mentioned at this point.

- **New network topology**: There have to be used new network topological systems which make it harder to detect them. Models like directed tree model or Gnutella network are mentioned at this point.

- **New communication mechanisms**: Also a new principle in topological communication behavior is necessary. Hackers will be wanting to prevent strong and striking data transfer rates between the infected host nodes. Probably a solution was, that every node was able to save information at itself, and that nodes keep usually communicating with direct neighbors only.

- **Content keying**: The communication between nodes could be keyed. Thinkable is also to hide such data inside of foreign files: for example in spam mails, in audio files (mp3) etc., for communicating purpose the todays file sharing networks (former Napster, eDonkey, eMule etc.) could be used. That example shows us, that not only computers inside of offices and firms could be a danger, also every private home user machine is concerned.

- **Home user computers**: Exactly these home user computers, who more and more get broadband Internet connection that stays permanently online, are a good target group for hackers. There are often computers that aren't safe enough in this user group. That goes farther, if there was used the possibility to send misinformation political information and similar things to produce some kind of information war.

- **Appearance changing**: Hiding the worm, using polymorphic and modular appearing increase the surviving chance for the worm.

- **Independent polymorphical change**: The worm should be able to change his face himself like an influenza virus. That could also be done with live updates that replace modules, change the behavior and more.

Some of these ideas have already become real and such worms are actually attacking computer systems [19].

5.10.3 Polymorphic traffic

There is foreseen that so-called zero-day exploits could be used. Zero-day exploits are security holes that are zero days old, in the ideal case. Commonly, you mean security holes that are still as good as unknown and also not fixed. This way - knowing about hardly-known issues - hackers are certainly getting a big and very dangerous advantage.

Multiple Vector attacks are another variant that is feared today. Multiple Vector means, that a worm is able to use more than only one possibility of spreading: attacks on web browsers, mail programs, chat clients, but also using special holes that were generated by other worms or viruses, and more. This kind of virus is extremely powerful in spreading and really hard to eliminate.

Also upcoming more and more are techniques to change the external appearance after breaking in a host. You could compare this action with the real, biological viruses - influenza for example - changing their shape some kind that our defensive system fails to recognize them. Also a computer virus can use this very mean trick to get unidentifiable and to hide in a very effective way.

There is already existing the possibility to separate special attacking code to single code fragments or also hazardously generated code signatures, so that they are not recognized by virus scanners. This trick is called polymorphic traffic. Machines specialized recognizing virus code by common string comparing will be failing. There are many possibilities you can separate a word to his ASCII code, or use some other key to translate it. Generating a signature matching engine that will compare and check every possibility at all is hardly possible [19].

5.10.4 Web crawlers as worms

Search engines in the Internet (Google, Altavista etc.) are using so-called web crawlers, automatically working bots who are searching and indexing the Internet, following every link they find in the normal case.

Hacker could use these crawlers for activating or generating some kind of damaging code. For example, there could be a possibility to generate a link that launches some e-mails, started by the web crawler testing it. Or some bad program files, that already lie on a reachable host, could be started by the spider calling that URL. There can also be generated damage on web pages, calling malicious CGI scripts on their hosts [19].

5.10.5 Jumping executable worm

Parent node attacks other servers and computer systems and and is spreading itself to more machines from there. But: as soon as a further system has been infected, the worm stops activity on the attack base system. The worm will only be active on one host at a time. This is a great advantage, because there is much less data transfer generated, no more activity on still infected hosts, and this way the worm is very much harder to be detected by human or automatic monitoring. It keeps scanning from only one host at a time and keeps his data inquirer's always under a special data sending rate threshold, to not be noticed. Normally there are always many other scans over the Internet on every host, it's very hard to find such a worm that behaves this way.

This technique also has a big disadvantage for the worm: if it really gets disturbed exactly on the one host where it is active (detection, host shutdown, connection problems etc.), the worm is as good as dead. It has lost his only active part, there is no spreading possibility anymore, because on all other infected hosts it keeps inactive.

Worms like this are not used for fast spreading jobs, but more special tasks like data theft in government organizations [19].

5.11 Computer security trends

5.11.1 Trustworthy computing philosophy (Microsoft)

In their own statements, they promise a philosophy for more security and trust. Four main points are important: security, data protection, reliability, integrity. Trustworthy computing was mentioned already in 2002. Experts and a lot of people criticize this idea: there were, and are, many severe security holes in Windows XP that are not repaired today, or the repairing was during too long time [20]. Experts are in doubt if Microsoft will manage this. Today, at the end of 2004, you can say that Windows XP is still bothered by security issues permanently. It seems to be very hard for a firm, if not totally impossible, to create a perfect safe operating system that covers about 90 percent of the

personal computer market. Additionally there are several fears about computer safety techniques named TCPA, and perhaps they are legitimated.

5.11.2 Windows and alternatives

There are several alternatives for Windows. Mac OS and Unix/Linux are mentioned mostly. Today, several countries, governments, cities, administrative institutions etc. are thinking about changing from Windows to Linux and open source software. An often heard argument: they don't want to be dependent on one firm (Microsoft). If this development is continuing in the future, there is a possibility that Microsoft looses some parts of market share to Linux and other alternatives. We would get a market that is more mixed up as today, with no mono culture. That would be a problem for hackers, because they will have to adapt their attacks to several operating systems and not just one. Today it's enough if a virus, worm etc. just attacks Windows computers, it can already spread out around the whole world and cause giant damage (see Code Red, I love you etc.). With a market that is more mixed up, there are much better chances that malicious ware will be stopped spreading earlier. That is a logic and natural fact that can be compared, for example, with agriculture. If you have mono cultures, there is a much bigger risk that just one kind of pest destroys your whole farm, as if you have a mixed farm.

What's the OS situation today? We looked at W3Schools website, an Internet organization that also counts web page visitors, identifying browser, operating system and more. You can see there that non-Microsoft operating systems are slightly, but continuously growing. In autumn 2004 till now, Linux counts 3.1%, Mac OS 2.6 to 2.7% of all visitors. Last year at the same time, both OS had about 0.5% less. But also Windows XP is growing, substituting Windows 2000 and other, old Windows versions. All versions together, Windows is still the market leader, installed on about, or even more than, 90% of all visitors computers. But it seems that a minority of computer users, who updates from an old Windows version, decides for an alternative OS instead of Windows XP.

5.11.3 Web browsers

Years ago, there was the so-called browser war between Netscape and Internet Explorer (IE). The Browser war went over, IE won because of much better quality in HTML and CSS rendering compared to Netscape 4. At that time, Netscape was not able to finish their next generation browser Netscape 5 fast enough (that was, by the way, the reason why Netscape jumped version 5 and directly went to version 6, preventing a deeper version number compared to IE that time). And concerning the IE, browser safety and specific security issues weren't that critical like today.

In 2004, the browser war is over. Or, perhaps better, a new browser war is going to start up. Short time ago, the Mozilla based open source browser Firefox 1.0 was released and seems to have a lot of advantages, compared to IE which is permanently fighting against security holes. Last thing hardly exists in Mozilla and Firefox, and if, such mistakes are repaired quickly. The W3Schools browser statistics (see figure 5.12) show that this new

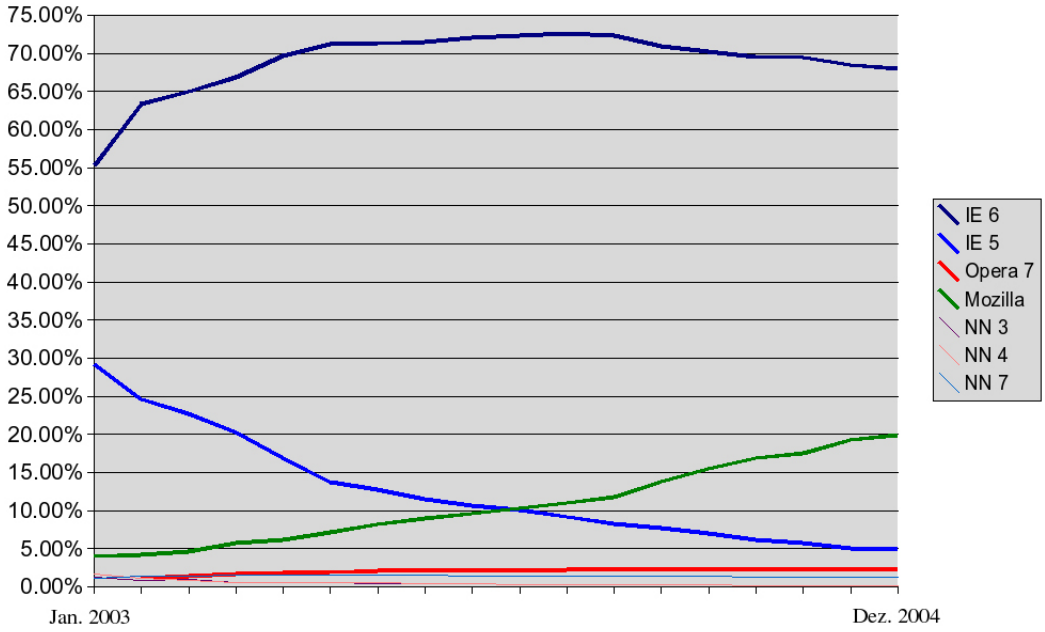

Figure 5.12: Browsers on the market in the last two years (Source: w3schools.com, own representation)

browser is strongly growing, especially there the Mozilla based browsers are heading to one fourth of all browsers (in March 2004 they had still less than 10%, in May 2003 less than 5%). The 20% border has been passed at the end of 2004, and this number is still strongly growing at a speed of about 1-2% per month. But be careful: this numbers are only useful to show the growing process, the absolute value may not be representative for the whole Internet browser scene. They have only be counted on w3schools.com with a more technically interested public as normal, compared to other websites there may be differences. Especially concerning the big Mozilla share: on other websites, with another kind of visitors, there may still IE be dominating with more than 90%. But the growing of Firefox importance seems to be a fact.

Possibly Firefox will still grow, especially at home user computers. This would be good for security, especially against DDoS attacks: it is getting much harder to find enough end user machines with specific security holes that can be attacked by a worm. There have to be developed techniques for both IE and Firefox in the minimum, if there should be enough infected computers.

5.12 Conclusion and outlook: possible financial consequences for enterprises

You can see some kind of problems and perhaps very big challenges are coming for enterprises with Internet applications (see figure 5.13). Top problems in the last years, and quite surely staying in the future, are the budget constraints. But also human resources lacks can be a problem, availability of skilled staff that was very important. Additionally

the staff has to learn every day, virus and security development will never stop and the hackers won't sleep, they are very intelligent and thinking opponents who have the possibility to be always one step farther. Very important is also the users security awareness, in the firms but also at home. Every computer system connected to the Internet could be a base for attacks. Special information security lessons for users working in the firms should probably be realized, but that is again a budget resource problem. There will be a hard way between investing as little money as possible for security and getting the best security as possible, also calculating what will happen if the biggest possible disaster gets real.

1994 Survey Results*	2003 Survey Results	2004 Survey Results
1. Lack of human resources	1. Budget constraints or limitations	1. Lack of security awareness by users
2. Budget constraints or limitations	2. Resource priorities	2. Budget constraints or limitations
3. Management awareness	3. Availability of skilled staff	3. Availability of skilled staff
4. Tools and solutions	4. Management commitment	4. Difficulty proving the value of information security
	5. Management awareness	5. Pace of information technology change

* *Only four obstacles were designated for this year's survey results*

Figure 5.13: Top Obstacles to Effective Information Security (Source: Ernst & Young (2004))

To prevent any attacks and accidents will probably be much more important than repairing damages. If customers are insured (and the law can require that), enterprises may also have to pay for possibly giant customer damages caused by malicious attacks. This financial risk is so high and is getting even higher, so enterprises have to do everything to prevent any bad influences by worms and similar attacks. One black day could be the end, otherwise.

Enterprises budgets will have to maintain the highest possible security level for all their computer and server infrastructure. Saving money here would be a bad idea. It has to be paid for security infrastructure like hardware and software, but also for security experts maintaining and monitoring the whole IT equipment every day.

There has also to be invested money for efficient data cryptographies, especially for data transfers.

What's about the todays really existing safety budget plans? We searched the Internet and found some interesting results. Ernst & Young Assurance and Advisory Business Services launched an enterprise survey in 2004 about IT security in firms, now the report is available. At a glance, the got the following facts:

- the majority does not give the information security a CEO level priority.
- the lack of security awareness by users is considered as the biggest problem of effective information security. Although, only a minority is investing in employee security training programs.

- the real security problem is by far more seen in malicious software, as in employee misconducting involving information systems.
- employee security and controls training is provided by less than the half of the responding enterprises.
- only a quite little part (about ten percent) thinks that government regulations would improve their IT security or reduce their data protection risks.

Firms know that there has to be more safety, but the safety must not cost more money [22].

And the general question: to prevent or to repair? Over all, we think the answer should better be to prevent. But that's often not as easy as you could think. The problem is, that security is a very expensive thing. And a 100% protection will never be reached at all. And: the closer you come to the 100% security, the faster the costs will rise for that. Enterprises will always have to find the best way between too little security and too high costs or the best compromise out of not too expensive on the one side, and as safe as possible on the other.

Bibliography

[1] Brian Shea: Have you locked the Castle Gate? Addison-Wesley (2003)

[2] John Canavan: Fundamentals of Network Security. Artech House (2001)

[3] Brian Bagnall, Chris O. Broomes, Ryan Russell: E-Mail Virus Protection Handbook. Syngress Publishing (2000)

[4] Ryan Russell, Teri Bidwell, Oliver Steudler, Robin Walshaw: Hack proofing your e-Commerce Site. Syngress Publishing (2001)

[5] AntiVirusLab. www.antiviruslab.com (Access: Nov 2004)

[6] Distrubuted Denial of Service (DDoS) Attacks/Tools. www.staff.washington.edu/dittrich/misc/ddos (Access: Nov 2004)

[7] Jelena Mirkovic, Janice Martin, Peter Reiher: A Taxonomy of DDoS Attacks and DDoS Defense Mechanisms. www.lasr.cs.ucla.edu/ddos/ucla_tech_report_020018.pdf (Access: Nov 2004)

[8] Bundesamt für Sicherheit in der Informationstechnik (BSI): Angriffstools - DDOS, Analyse der Angriffstools. www.bsi.de/fachthem/sinet/literatddos/toolsana.htm (Access: Oct 2004)

[9] eCrime Watch Summary 2004. www.cert.org/archive/pdf/2004eCrimeWatchSummary.pdf (Access: Oct 2004)

[10] NHTCU Hi-Tech Crime Survey for 2003. www.nhtcu.org (Access: Dec 2004)

[11] Edmund E. Lindau: DDoS-Schutzgelderpressung erstmals auch in Österreich. www2.computerwelt.at/detailArticle.asp?a=80670&n=4 (Access: Oct 2004)

[12] Bundesamt für Sicherheit in der Informationstechnik (BSI) (Dr. Stefan Wolf, Dr. Dirk Häger, Heiner Schorn): Angriffe - Erkennung und Behandlung von Angriffen aus dem Internet. www.bsi.de/fachthem/sinet/webserv/angriff.htm (Access: Oct 2004)

[13] Easyjet. www.easyjet.de/DE/Unsere/Informationspaket (Access: Oct 2004)

[14] Computerbetrug.de. www.Computerbetrug.de (Access: Oct 2004)

[15] Heise online (Holger Bruns): Ultimatum soll Spammer unter Druck setzen. www.heise.de/newsticker/meldung/26232 (Access: Nov 2004)

[16] Bundesamt für Sicherheit in der Informationstechnik (BSI): Schutz vor DDoS Empfehlungen zum Schutz vor verteilten Denial of Service-Angriffen im Internet. www.bsi.de/fachthem/sinet/ddos.htm (Access: Oct 2004)

[17] Ernst & Young: Globally security piece of information program 2004

[18] toplayer.com. www.toplayer.com/content/products/intrusion_detection/attack_mitigator.jsp (Access: Nov 2004)

[19] Jose Nazario: Defense and detection strategies against Internet worms. Artech House (2004)

[20] tecchannel.de: Experten misstrauen Trustworthy Computing. www.tecchannel.de/news/20020430/thema20020430-7425.html (Access: Nov 2004)

[21] Mummert Consulting AG: Mangelnder Verbraucherschutz bei Online-Finanzdienstleistern. www.ecommerce-trends.de/0303_02.htm (Access: Nov 2004)

[22] derStandard.at: IT-Sicherheit wird immer wichtiger, darf aber nicht mehr kosten. www.derstandard.at/?url=/?id=1845788 (Access: Nov 2004)

[23] Ernst & Young: Global Information Security Survey 2004

[24] Microsoft: Trustworthy computing. www.microsoft.com/germany/ms/security/twc/default.mspx (Access: Nov 2004)

[25] W3Schools: Browser Statistics. www.w3schools.com/browsers/browsers_stats.asp (Access: Nov 2004)

Kapitel 6

Das Individuum als Operator

Ante Bujas, Christian Breuel, Manuel Ziltener

Diese Seminararbeit wirft grob umfasst einen verallgemeinerten Blick in die heutige Telekommunikationslandschaft und in die verwendete Telekommunikationstechnik, und zeigt, inwiefern sich die heutige Operatorenlandschaft von ihrer Vergangenheit unterscheidet und was die wichtigsten Ursachen dafür sind. Die verschiedenen Überlebensstrategien der zurzeit dominanten Marktteilnehmer werden angschaut, und einige rechtlichen Aspekte werden aufgeworfen und hinterfragt. Dies in subtiler aber durchaus gerechtfertigter Weise. Des Weiteren werden disruptive Technologien angeschnitten, es wird erklärt, was diese darstellen und wieso sie in gewissen Situationen namentlich disruptiv wirken. Am Schluss wird noch vorsichtig aber überlegt ein Blick in die Nahe und Ferne Zukunft geworfen, die erwartungsgemäss viele Chancen aufzeigt aber auch nicht zu unterschätzende Risiken verbirgt. Um diese Seminararbeit in durchwegs adäquatem Umfang zu halten und um die wichtigsten Aspekte des Themas eindeutig aufzuzeigen, wird primär die Wireless Local Area Network (WLAN) Technik aufgegriffen, mit der sich die verschiedenen Thematiken des Themas gezielt und verständlich aufzeigen lassen. Keineswegs beschränkt sich die tiefe Thematik unsers Themas nur auf diese eine Technologie, sie umhüllt vielmehr ein deutlich grösseres Spektrum, welches in sich jedoch viel zu heterogen und daher unübersichtlich ist, um die einzelnen Akzente des Themas geschickt aufzugreifen und zu erklären.

Inhaltsverzeichnis

6.1 Einführung

Der Mensch als Individuum in einer in sich heterogenen Gesellschaft versucht schon seit eingier Zeit die Stellung eines Operators einzunehmen. Diese Versuchung ist umso stärker, je mehr sich die unmittelbar betroffene Umgebung davon zu schützen versucht. Die Frage ist, was überhaupt ein Operator darstellt und wieso ein Individuum sich von dieser Stellung so angezogen fühlt. Auch wenn die Funktion als Operator viele Vorteile mit sich bringt, so bringt sie auch nicht zu unterschätzende Nachteile mit sich. Diese inhärente Diskrepanz ist durchwegs wünschenswert, da sie einen Wildwuchs in der Operatorenlandschaft implizit verringert, was die Attraktivität dieser Landschaft langfristig und nachhaltig erhöht. Ein Widerspruch in sich ist dies nicht, da weder die Operatoren noch die Operatorenlandschaft sich was anderes wünschen. Auch wenn sich die verschiedenen Operatoren mehr Handlungsspielraum wünschen, so empfinden sie diese explizite Einschränkung durchaus auch als Vorteil.

Der heutige Telekommunikationsmarkt wird von einzelnen grossen Unternehmen dominiert. Diese Unternehmen wirtschaften in einem Markt, der sehr umkämpft ist. Innere aber auch äussere Kräfte bedrängen diesen rund um die Uhr, und versuchen die Macht und den Einfluss der grossen und etablierten Marktteilnehmer wenn möglich nachhaltig zu dämpfen. Diese Kräfte kommen von direkten Konkurrenten aber auch von anderen Einflüssen wie Regulierungen und Gesetzen. Gerade gewisse einschneidende Regulierungen sind trotz den anhaltenden Deregulierungsrufen beziehungsweise Tendenzen immer noch vorhanden und sind für den gesamten Telekommunikationsmarkt, und für die betroffenen Unternehmen sowieso, meist nachhaltig und langfristig ein nicht zu unterschätzendes Hindernis.

Betrachtet man den Telekommunikationsmarkt der Schweiz, so ist zu erkennen, dass wenige Unternehmen den grössten Teil des Marktes abdecken und die kleineren Unternehmen sich das gebliebene Reststück untereinander aufteilen müssen. In der Schweiz ist der dominanteste Anbieter unübersehbar die Swisscom, deren Monopolstellung eindeutig und unangefochten ist. Swisscom hatte zum Beispiel im Jahre 2001 im Mobilfunkmarkt einen Marktanteil von 63.8 Prozent[1]. Die Swisscom konnte bis heute ihre Marktführerschaft ganz erfolgreich verteidigen.

Auch wenn das Marktumfeld allgemein in den letzten Jahren aggressiver und herausfordernder wurde, die Swisscom war und ist weiterhin die Nummer eins. Neben der Swisscom ist noch die Cablecom zu erwähnen, die auch sehr erfolgreich im Telekommunikationsmarkt operiert. Die Cablecom ist der erfolgreichste Kabelnetzanbieter in der Schweiz[2] und ist erst vor einigen Jahren ins eigentliche Internetgeschäft eingestiegen und konnte sich bis jetzt entgegen vielen oft pessimistischen Meinungen gut im Markt bewähren. Neben diesen zwei aufgezeigten Marktführern gibt es selbstverständlich noch andere, kleinere Unternehmen, die ganz oder nur teilweise im Telekommunikationsmarkt interagieren oder nur einen kleinen Teil dessen abdecken. So zum Beispiel das Unternehmen Bluewin[3], welches primär als Internet Service Provider auftritt oder Sunrise[4], die im Bereich der Mobiltelefonie stark vertreten ist und auch weiterhin bleiben wird.

Das Angebots-Spektrum ist sehr breit und durchmischt gefächert, und geht von reinen Internet Service Providern bis zu Unternehmen, die alle Telekommunikations-Services und

Dienste anbieten. Die Portfolios sind sehr heterogen und von Unternehmen zu Unternehmen verschieden diversifiziert und ausbalanciert. Klar stellt sich da automatisch die Frage nach der Überlebensfähigkeit einzelner Anbieter, welche bewusst, aber teilweise sicher auch ungewollt und zwangsläufig ein auffallend kleines Portfolio aufweisen. Die Antwort ist nicht einfach zu finden. Auch wenn die Konkurrenzfähigkeit einzelner kleineren Unternehmen oft in Frage gestellt wird, so sind es gerade diese Firmen, die neue Technologien oder neue Dienste anbieten und so immer wieder mit Produktinnovationen den Markt überraschen. Doch reine Produktinnovationen sind nicht ausreichend um sich im Markt langfristig und nachhaltig zu profilieren, vielmehr sind gut durchdachte Marketing- und Angebotsstrategien nötig, und gewiss auch Qualitäts- und Produktverbesserungen.

Unter den verschiedenen breit gefächerten Strategien der verschiedenen Telekommunikationsunternehmen sind nicht überraschend einige immer wiederkehrende Strategie-Muster zu erkennen. Hier einige Beispiele:

- *Erweitertes Produkte Angebot:* Die Unternehmen erweitern gezielt ihr Portfolio und wollen neue Kunden mit neuen Produkten gewinnen. Zum Beispiel UMTS[5]
- *Verbessertes Angebot:* Vorhandene Produkte werden bezüglich Qualität und Ausstattung stetig verbessert. Zum Beispiel höhere Bandbreiten bei Modems[6] und Routers, mehr Anschlüsse bei den Routers und Switches, besseres und einfach zu Bedienende Konfigurationsmenu des Modems oder Routers.
- *Verbilligte Angebote:* Die Produkte und Dienste werden immer billiger angeboten[7], wobei hier zu sagen ist, dass gerade in der Schweiz einige Dienste wie die Mobiltelefonie noch zu teuer ist.
- *Vereinfachter Einstieg für Kunden:* Viele Unternehmen vereinfachen den Einstieg indem sie zum Beispiel die Aufschaltgebühren bei einem ADSL Anschluss weglassen oder das ADSL-Modem gratis anbieten[8]

Fazit:

Die grossen Anbieter tun alles Mögliche um ihre Marktposition zu verteidigen, aber auch um ihre Angebote und deren Qualität zu verbessern.

6.2 Die Technik

Die Technik, insbesondere die Computertechnik hat sich in den letzten Jahren und sogar Jahrzehnten enorm weiterentwickelt. Was früher noch immens teuer und exklusiv war, ist heute Standardausrüstung.

Hier einige Beispiele:

- Bei den heutigen Laptops sind die WLAN Karte, die Netzwerkkarte, das Bluetooth und noch andere Technologien von Grund auf integriert, es sind also keine oder nur wenige Aufrüstungen nötig. Zudem ist die Hardware schon vollständig installiert und daher vollumfänglich funktionstüchtig.
- Die Modems, Routers, Switches und die Wirelesskarten sind mittlerweile sehr billig und werden immer billiger. Dieser Umstand ist ein grosser wenn nicht sogar der grösste Vorteil.
- Auch die Access Points werden immer billiger und vor allem immer leistungsfähiger. Eine flächendeckende Überdeckung mit WLAN ist mittlerweile kein Problem mehr. Radien von 50 Metern und mehr sind keine Seltenheit mehr. Zudem ist die Empfangsqualität entscheidend besser geworden, dies nicht zuletzt wegen der fort schreitenden Technik.
- Modem, Router und Access Point sind immer öfters zusammen integriert Der Kunde muss sich nur ein Gerät kaufen und hat alles nötig mit dabei. Die Handhabung und Installation der nötigen Infrastruktur ist einfacher geworden.

Die Technik aber auch das Angebot hat sich in den letzten Jahren fundamental verändert. Der Kunde, die Abnehmer, der Konsument kann von dieser Entwicklung vollumfänglich profitieren. Der finanzielle aber auch physische Aufwand ist für ihn wesentlich gesunken. Die Installation eines Internetzugangs wird immer einfacher was den Internetboom noch zusätzlich fördert.

Es wird aber nicht nur einfacher und billiger zu Hause ein Netzwerk zu installieren sondern auch einfacher, seiner unmittelbaren Umgebung dieses Netzwerk zur Verfügung zu stellen. Gerade diese Tatsache trifft den Kern dieses Themas. Der Einzelne, der Mensch, das Kleinunternehmen, das Grossunternehmen oder die Familie XY kann von nun an sehr einfach und billig ein eigenes Netzwerk mit Internetzugang aufbauen und diesen auch eigenständig verwalten. Das Verwalten von Netzwerken war lange Zeit ausschliesslich in den Händen von etablierten Telekommunikationsunternehmen. Nun liegt es auch in den Händen des Einzelnen, des Individuums.

6.2.1 Beispiele verschiedener Operationsmodelle

Um dieses Phänomen deutlich zu machen, folgen nun einige Beispiele, die im ersten Moment zwar als nicht aussergewöhnlich oder sogar trivial erscheinen mögen, wegen ihrer

Einfachheit und Gewohntheit jedoch sehr repräsentativ die Thematik Individuum - Operator aufzeigen.

Die ersten zwei Beispiele sind von nicht gewerblicher Natur. Sie zeigen, inwiefern sich Individuen als Operatoren etablieren können und inwieweit die unmittelbare Umgebung davon profitieren kann. Beispiel drei und vier nehmen zusätzlich noch den wichtigen Aspekt des Geschäftsmodells unter die Lupe. Sie zeigen, dass das Individuum als Operator auch gewerblich profitiert und dass er seine Operatorenstellung zur Erweiterung seines Portfolios nutzen kann und somit generell für seinen Kunden auch einen Mehrwert generiert.

Beispiel 1: Mehrfamilienhaus

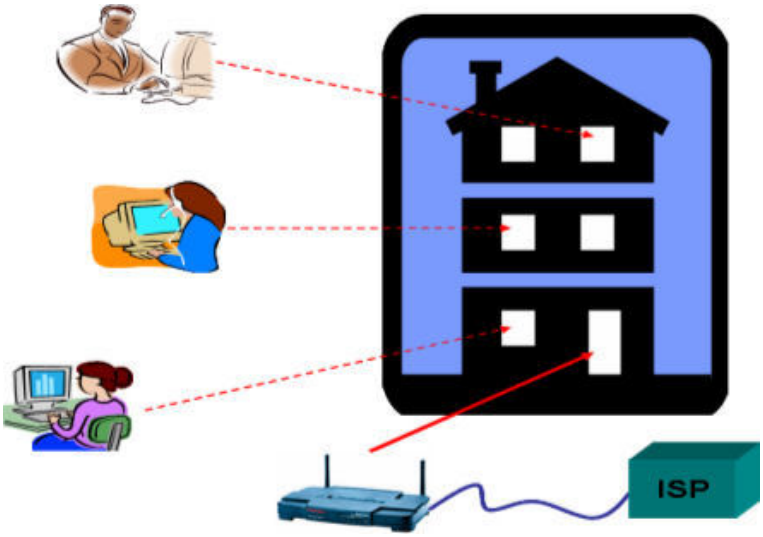

Abbildung 6.1: Mehrfamilienhaus

Ein Haus mit mehreren unabhängigen Familien wie in Abbildung 6.1 dargestellt. Nur ein Internetzugang im ganzen Haus. Alle Mitbewohner haben uneingeschränkten Zugang ins Internet. Die Kosten, die für den Internetzugang anfallen, werden unter den Mitbewohnern beziehungsweise Familien aufgeteilt. Das ist auch der Hauptgrund. Des Weiteren muss nur eine Person oder Familie oder sogar einfach nur der Hauswart von der Technik, der Software und deren Handhabung informiert sein. Es wird also nicht nur der finanzielle Aufwand verteilt, sondern auch der übrige Aufwand. Der Nachteil ist die Zentralisation der Infrastruktur und des Wissens. Entstehen zum Beispiel technische Störungen mit dem einzigen Modem, so sind alle Mitbenutzer davon betroffen. Das Problem verschärft sich noch mit der möglichen Tatsache, dass das nötige technische Wissen über die Handhabung und Behebung von Störungen bei nur einer einzigen Person liegen kann.

Beispiel 2: Die Community

Jeder zahlt und installiert einen Anschluss, teilt ihn aber auch mit anderen und kann so auch andere Netzwerke benutzen. Dies wird in Abbildung. 6.2 weiter verdeutlicht.

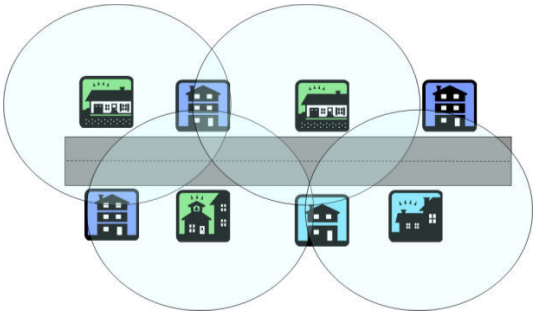

Abbildung 6.2: Community

Eine wichtige Voraussetzung ist in diesem Fall eine Vereinheitlichung der benutzen Technik, was einer Standardisierung sehr nahe kommt. Damit dieses Konzept funktioniert müssen sich die Beteiligten Netzwerkverwalter auf eine einheitliche Hardware und Software einigen. Nur so kann ein problemloser Zugriff auf das Netzwerk garantiert werden

Beispiel 3: Das Restaurant

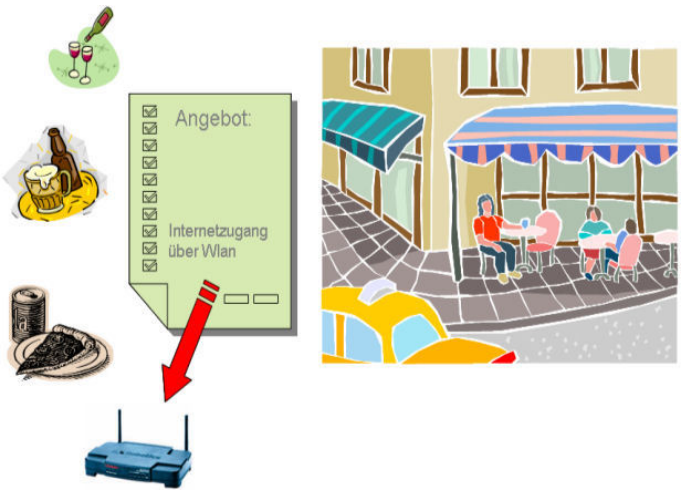

Abbildung 6.3: Restaurant

Der Besitzer eines Restaurants oder Pubs will seinen Kunden einen Mehrwert bieten indem er einen Internetzugang über WLAN anbietet. In Abbildung. 6.3 wird dieses Geschäft weiter verdeutlicht. Dadurch erhofft er sich mehr Einnahmen zu generieren. Dies erreicht er durch zufriedenere und allgemeinen mehr Kunden. Der Internetzugang ist also Teil seines Geschäftsmodells.

Beispiel 4: Wireless-Lan Zonen

Eine Stadt/Gemeinde bietet WLAN an. Dadurch erhofft sie sich unter anderem:

- Höhere Attraktivität
- Mehr Touristen
- Mehr Geschäftsleute
- Höhere Bekanntheit

6.2.2 Unterschiede in den Geschäftsmodellen

Das Beispiel mit dem Restaurantbesitzer im Beispiel drei zeigt explizit, wie ein Internetzugang in ein Geschäftsmodell integriert werden kann. Durch die Erweiterung des Angebotes

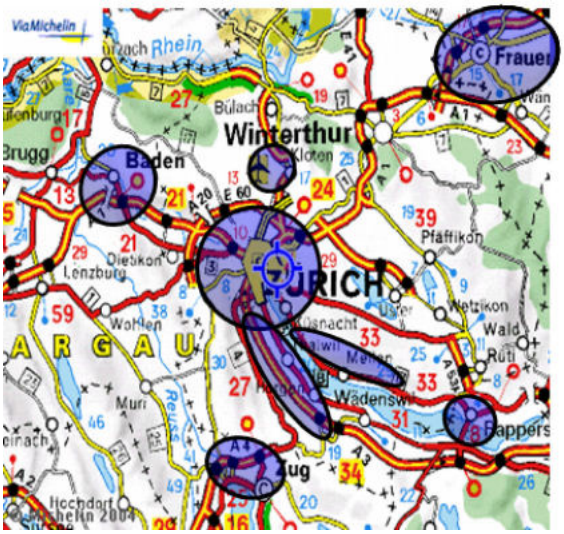

Abbildung 6.4: Vision

Die Stadt/Gemeinde will primär ihre Bekanntheit erhöhen, indem sie einen zusätzlichen Service bietet. In der Karte in Abbildung 6.4 wird die Netzabdeckung dieser Services verdeutlicht. Dieses Szenario entspricht nicht der Gegenwart sondern widerspiegelt viel mehr eine mögliche Zukunftsvision. Die Wahrscheinlichkeit dass eine solche Vision in naher oder ferner Zukunft wahr wird, ist ziemlich hoch.

um einen Internetzugang sollen mehr Kunden angesprochen werden. Der Unternehmer erhofft sich nicht nur neue Kunden, sondern auch zufriedenere Kunden, die immer wieder gerne in sein Geschäft beziehungsweise in sein Restaurant kommen. Die positive Resonanz der Kunden ist ein Garant für den Erfolg. Der dadurch generierte Mehrwert beeinflusst den Unternehmensgewinn positiv und ist nicht zuletzt wesentlich verantwortlich für neue Investitionen. Gerade in der heutigen Zeit, in der die Computer Technik sehr schnell fortschreitet und sich die Allgemeinheit immer mehr dieser aufdringlichen Technik unterwerfen muss, sind solche Geschäftsstrategien für einen soliden Erfolg gerade prädestiniert.

Solche Geschäftsmodelle, wie oben beschrieben, sind heutzutage schon oft anzutreffen. Auch wenn sie von Auge schwer zu sehen sind, so sind sie schon weitläufig vorhanden. Die Restaurantkette McDonald's zum Beispiel bietet schon seit längerem einen Internetzugang über WLAN in ihren Restaurants an. Wer im Besitz eines Netzwerkfähigen Laptops ist kann während seiner Essenszeit genüsslich im World Wide Web rumsurfen. Wichtig ist hier zu wissen, dass McDonald's das WLAN und die ganze Infrastruktur nicht selber betreibt und unterhält, sondern ein externes Unternehmen dies tut. Siehe Grafik[9]

Das externe Unternehmen ist voll verantwortlich für die Installation, die Instandhaltung und Wartung der WLAN Infrastruktur, aber auch für die Qualität der von ihnen angebotenen Dienstleistungen. Eines dieser Unternehmen ist Monzoon: siehe Abbildung 6.5.

Ein weiteres Beispiel bezüglich angebotenen Internetzugangs sind diverse Hotels, die in ihren Aufenthaltsräumen ebenfalls WLAN anbieten. Auch da sind externe Firmen mehrheitlich involviert[10]. Die Installation, Instandhaltung und Wartung übernimmt in diesem Fall die Swisscom. Neben diesen Geschäftsmodellen, in denen externe IT-Firmen verantwortlich für die ganze WLAN Infrastruktur sind, gibt es erwartungsgemäß auch solche Modelle, in denen die Geschäftsbesitzer gleichzeitig auch die WLAN Ausrüster und Betreiber sind. Der Geschäftbesitzer ist vollumfänglich für den Betrieb seines Netzwerkes verantwortlich. In diesem Fall ist eine Unterscheidung von großer Bedeutung:

- Der Geschäftsbesitzer betreibt seine eigenen Internetzugang gewerblich, daher, die

Abbildung 6.5: Monzoon

Kunden müssen zusätzlich für diese Dienstleistung zahlen. Der Kunde bekommt den Internetzugang nicht implizit durch andere Geschäftstätigkeiten. Der Internetzugang ist für sich ein eigenständiges Produkt neben den anderen Produkten

- Der Geschäftsbesitzer betreibt seinen eigenen Internetzugang nicht gewerblich. Er bietet in so zu sagen gratis an. Die Kunden müssen nicht explizit dafür zahlen sondern erhalten diesen Dienst als Beilage einer anderen Geschäftstätigkeit. Die Restaurantbesucher bekommen also zum Beispiel einen Internetzugang bei der Konsumation eines Getränkes.

6.2.3 Probleme und Fragen

Wie in den vorhergehenden Beispielen ersichtlich wurde, wird die Rolle des Operators immer mehr eingenommen. Dieser Rollentausch ist zwar gut gemeint, gleichzeitig führt diese Entwicklung aber sehr schnell zu einem unerwünschten Wildwuchs in der Operatorenlandschaft, welcher langfristig sehr gefährlich und unkontrollierbar ist. Die erhofften Vorteile sind insofern trügerisch, als dass sie zwar der Allgemeinheit einen Mehrwert offerieren, gleichzeitig aber auch einen nicht gleich ersichtlichen Tribut von ihnen fordern. Dieser Tribut ist keineswegs von kleinem Wert und noch weniger von trivialer Natur, sondern schneidet großflächig die Identität und Persönlichkeit jedes Einzelnen an. Was sich auf den ersten Blick als kundenfreundliche und daher wertvermehrende Geschäftsidee etabliert, kann sich auch als eine nicht zu Ende durchdachte Geschäftsstrategie entpuppen. Dies ist jedoch nicht immer so. Folgend werden einige Problematiken kurz angeschnitten, die Antworten zu einzelnen Fragen werden im nächsten Abschnitt beantwortet.

Sicherheit

Gerade die grossen Telekommunikationsanbieter tun alles Mögliche um persönlichen Daten ihrer Kunden zu schützen. Diese Problematik trifft die neuen kleinen Operatoren umso mehr, da sie weder Erfahrungen noch die nötigen finanziellen Mittel haben. Es liegt sicher auch im Interesse des Geschäftbesitzers die Kundendaten ausreichend zu schützen.

Rechtliche Probleme

Darf ein Restaurantbesitzer einfach so einen Internetzugang anbieten oder muss er gewisse Regeln und Pflichten einhalten? Wer haftet im Falle eines Missbrauches seines Internetzuganges? Ist es erlaubt für einen Internetzugang gewerblich zu gebrauchen?

Qualität

Wie der Aspekt der Sicherheit so ist auch die Qualität ein entscheidender Faktor. Auch hier tun die etablierten Telekommunikationsanbieter alles Mögliche, um die Qualität ihrer Dienstleistungen stetig und nachhaltig zu verbessern. Auch hier stellt sich wiederum die Frage, ob die kleinen Operatoren für Qualität sorgen können, obwohl sie weder finanziell noch vom Wissen her den Grossanbietern die Stirn bieten können.

6.2.4 Fazit

Die Rivalität zwischen den etablierten Telekommunikationsanbieter und den neuen kleinen Operatoren nimmt stetig zu. Aus der Sicht der Grossanbieter ist dieser Umstand durchaus gerechtfertigt. Es waren, beziehungsweise sind ja gerade sie, die sich konsequent für mehr Sicherheit und verbesserte Qualität im Internetmarkt einsetzen. Die kleinen Operatoren sind aber durchaus auch motiviert dies zu tun, nur sind sie der Wissens- und Technikflut der heutigen Zeit hemmungslos ausgesetzt und dies verunmöglicht es ihnen im besonderen Masse, die hohen Sicherheits- und Qualitätsstandards zu erreichen. Mittel- und langfristig wird dies eine grosse Herausforderung für die kleinen Operatoren sein. Wer sie nicht meistern kann wird erbarmungslos zu Grunde gehen. Den grossen Telekommunikationsanbietern kann dies nur recht sein.

Die Operatoren, bezwiehungsweise die, die es mal werden wollen, sind an verschiedene rechtlichen und gesetzlichen Auflagen gebunden. Diese Grundlagen sind insofern von hoher Bedeutung, als dass sie explizit für eine überschaubare und organisierte Operatorenlandschaft sorgen. Das ist, wie schon angedeutet, auch im Interesse der Betroffenen, da auch sie rechtlich einwandfreie und qualitätsorientierte Internet-Services bieten wollen. Rechtliche Aspekte können jedoch ziemlich komplex sein, sind aber eben sehr wichtig. Der folgende Abschnitt geht daher genauer auf einige wichtige rechtliche Aspekte und Fragen ein.

6.3 Rechtliche Fragen

In diesem Teil wird auf rechtliche Fragen in Bezug der freien Nutzung von Zugängen zum Internet eingegangen. Zu Beginn werden die vom Bundesamt für Kommunikation aufgestellten Richtlinien besprochen und analysiert, was sie bedeuten. Es werden verschiedene Szenarien von Missbräuchen aufgezeigt und versucht eine rechtliche Grundlage dazu zu finden.

Der zweite Teil beschäftigt sich mit der Sicherheit von mobiler Datenkommunikation. Dabei sollen Stärken und Schwächen der einzelnen Verfahren aufgezeigt werden und auf deren Verwendungszweck kurz eingegangen werden.

6.3.1 Verschiedene rechtliche Fragen

Wird ein Anschluss und damit die Bandbreite unter mehreren Personen aufgeteilt, werden automatisch auch Fragen entstehen. Wer haftet, wer erhält wieviel Bandbreite, werden verschiedene Aufgaben in unterschiedliche Prioritäten eingestuft oder wer ist für die Sicherheit verantwortlich? Einige dieser Fragen werden durch das Gesetz geregelt aber andere müssten per Vertrag oder mündlichen Abmachungen geklärt werden.

Die rechtlichen Grundlagen über die Fernmeldedienste ist laut Bundesverfassung Art 92 Abs1 Sache des Bundes und wird im Fernmeldegesetz geregelt.

Meldungspflicht für Anbieter von Fernmeldediensten

Für das Anbieten eines Zuganges zum Fernmeldewesen, wozu die Datenübertragung zählt, verlangt das Bundesamt für Kommunikation BAKOM die Meldung. Zitat aus dem Faktenblatt für Radio Local Area Network des BAKOM [11]: Werden RLAN-Anlagen für das Betreiben eines Fernmeldenetzes verwendet, mit welchen eine Anbieterin für Dritte (Teilnehmer oder andere Fernmeldedienstanbieterinnen) Fernmeldedienste8 (z.B. Sprachübertragung, Datenübertragungsdienste,...) erbringt, so ist eine Dienstkonzession gemäss Art. 4 Abs. 1 FMG erforderlich. Dabei spielt es keine Rolle ob das RLAN für den Teilnehmeranschluss oder für die Vernetzung von Fernmeldeanlagen eingesetzt wird. Die einmaligen Kosten einer Konzession betragen laut den Verwaltungsgebühren im Fernmeldebereich [15] des Departements für Umwelt, Verkehr, Energie und Kommunikation UVEK zwischen 3'000 - 5'000 CHF. Dazu kommen jährlich wiederkehrende Kosten von 1000 CHF.

Damit unterschieden werden kann, welcher Anbieter von WLAN unter die Konzessionspflicht fällt, unterscheidet das BAKOM mittels folgender Definitionen zwischen privaten und öffentlichen WLAN's[12].

Privates WLAN

- Dieses ist reserviert für und betrieben durch eine bestimmte für das Lokalnetzwerk authentifizierte Benutzergruppe. Eventuell werden Besuchern temporär Gastdienste angeboten. In diese Kategorie fallen z.B. Heimnetzwerke oder Geschäfts-LANs.
- Ein WLAN an quasi öffentlicher Stelle, jedoch auf ein Gebäude oder Gelände beschränkt (Ausnahmebestimmung gemäss Art. 2 der Verordnung über Fernmeldedienste), das den Benutzern den Zugang zu einem öffentlichen Netz ermöglicht. Der WLAN-Betreiber hat entweder eine direkte Beziehung zum Teilnehmer oder z.B. über ein Roaming-Abkommen, eine indirekte Beziehung zum Teilnehmer.

Öffentliches WLAN

- Anbieter von WLAN-Diensten an öffentlichen Stellen: ein Abonnement-basierter Zugang zu Fernmeldediensten wird angeboten; Netzwerkanschlüsse werden mit WLAN zur Verfügung gestellt; die Nutzung wird registriert und verrechnet.
- Anbieter von WLAN-Infrastruktur an öffentlichen Stellen, wie z.B. Stadtzentren: der WLAN-Betreiber hat selber keine direkte Kundenbeziehung zu einem Teilnehmer. Die Nutzung der Infrastruktur wird im Wiederverkauf an einen bereits registrierten Fernmeldedienstanbieter angeboten.

Fernmeldedienstanbieter ist der Betreiber eines öffentlichen WLAN. Kein Fernmeldedienstanbieter ist der Betreiber eines privaten WLAN. Der uns interessierende Teil ist das öffentliche WLAN. Das öffentliche WLAN, wie es im Teil 2 der öffentlichen WLAN beschrieben wird, ist interessant für Kommunities, während der erste Teil mehr für die kommerzielle Nutzung dienlich scheint.

Ein für sich persönlich installiertes WLAN Netzwerk fällt also nicht in die Meldepflicht des BAKOMs, auch nicht wenn das WLAN für das ganze Gebäude freigegeben wird.

Der Preis, der für eine Fernmeldedienstkonzession anfällt (siehe oben), ist für Private eine nicht unterschätzbare Hürde. Es ist jedoch zu bezweifeln, ob eine Privatperson, welche ihr Wireless Lokal Area Network der Öffentlichkeit zur Verfügung stellt, über die rechtliche Situation informiert ist. Das Erbringen eines Fernmeldedienstes, wird nach dem Fernmeldegesetz 9. Kapitel Art. 52 Abs. 1a mit Haft oder Busse bis zu 100'000 CHF bestraft. Quelle Verwaltungsgebühren im Fernmeldebereich

Probleme, beschrieben an verschiedenen Szenarien

Durch die Digitalisierung von Informationen entstehen folgende Konsequenzen:

- Digitalisierung –> Jede Information ist in computerlesbaren Bits gespeichert
- Vulnerabilität –> (gesteigerte) Verletzbarkeit des geistigen Eigentums

- Alterabilität Digipulierbarkeit –> z.B. leichte Veränderbarkeit geistiger Werke
- Ubiquität –> Digitale Inhalte können überall abgefragt werden

[32]

Anonymisierung

Bei einer breiten Öffnung der Zugänge durch Private wird die Anonymisierung, der sich im Internet befindlichen Personen, vorangetrieben. Private sind nicht in der Lage oder haben auch keinen Ansporn, andere Teilnehmer zu authentifizieren, Daten über die mit ihnen verbundenen Benutzern zu sammeln, zu speichern und über eine gewisse Zeit aufzubewahren. Durch diese Praxis haben Hacker, Spammer und andere im illegalen Bereich operierende Personen leichtes Spiel ihre Spuren zu verwischen. Was in einem Internet Café durch die eingeschränkte Privatsphäre etwas eingeschränkt wird, ist beim eigenen Laptop oder sonstigen mobilen Einheiten weniger ein Problem. Spezifische Software kann auf diesen Units ungehindert installiert und über den anonymisierten Zugang ohne Einschränkungen verwendet werden.

Verantwortlichkeit

Ein weiteres Problem ist sicher die immer grössere Verbreitung von illegalen Inhalten im Internet. Kinderpornographie, rassistische Webseiten und sonstige obszöne und abstossende Inhalte können über offene private Zugänge anonym und damit straffrei konsumiert und verbreitet werden. Was geschieht nun, wenn über den Zugang oder besser über die die IP-Adresse eines Anbieters verschiedene illegale Inhalte konsumiert oder verbreitet werden? Kann der Anbieter in so einem Fall dafür haftbar gemacht werden?

Nehmen wir an, der Anbieter weiss nicht einmal, dass sein Zugang durch Dritte benutzt wird. Er hat mit dem Breitbandanschluss auch gleich einen Wireless Accesspoint erhalten und dieser nach Anleitung installiert. Es wurden bei der Installation jedoch keinerlei Sicherheitsvorkehrungen getroffen. Ein Dritter, der über Kenntnisse verfügt, kann das fehlen von Sicherheitsmassnahmen ausnützen und seinen jenen Anschluss missbrauchen. Kann der nichtwissende Anbieter dennoch belangt werden?

Die gesetzlichen Grundlagen bezüglich Computer- und Internetkriminalität sind in folgenden Artikeln geregelt (Tabelle. 6.1):

Missbrauch

Ein weiteres Diskussionsthema ist der Verlust oder der Diebstahl von Daten, während der Benutzung eines angebotenen Zugangs. Wer haftet, wenn ein Benutzer sich über einen angebotenen Zugang verbindet, und dieser Zugang so unsicher ist, dass seine Daten einfach abgehört werden können oder er durch Sicherheitslücken die Kontrolle über seinen

Tabelle 6.1: Gesetzesartikel

Coputerspezifische Straftatsbestände	Medieninhaltsdelikte	übriges Strafrecht
StGB 143 StGB 143^{bis} StGB 144^{bis} StGB 147 StGB 148 StGB 150	StGB 173 StGB 276	StGB 135 StGB 197 (Ziff3) StGB 261^{bis} URG 67

Computer verliert? Wieviel Wert ist demnach eine sichere Datenverbindung? Diese Frage lässt sich sicher nur mit der Diversifikation der Daten beantworten. Die Art der Daten, die man abrufen oder übertragen will, spielt dabei eine erhebliche Rolle. Für einen kurzen Blick auf Wetterinformationen oder Datenaustausch für Unterhaltungszwecke reicht wahrscheinlich auch ein etwas unsicherer Zugang. Will man aber sensible Daten, wie zum Beispiel ein Telefonat mit Voice over IP oder wichtige eMail oder vertrauliche Dokumente herunterladen, so kann der Zugang gar nicht sicher genug sein. Solange noch viele Benutzer eigentliche „Laien“ sind im Umgang mit Computer, im Speziellen mit sicherheitsspezifischem Verhalten, solange kann man nicht erwarten, dass ein Benutzer selber Kenntnis besitzt, wie man eine sichere Datenverbindung aufbaut.

Nicht nur rechtliche Aspekte sind von grosser Bedeutung. Auch Sicherheitsaspekte dominieren das Problemfeld und sind nicht weniger Wichtig. Ihre Komplexität kann in manchen Fällen sogar die Komplexität der rechtlichen Aspekte übertreffen. Im nächesten Abschnitt werden sicherheitsrelevante Probleme angschnitten.

6.4 Sicherheitsmassnahmen bei mobiler Datenkommunikation

Die Sicherheitsbedürfnisse werden mit der Mobilität immer mehr ansteigen. Die Benutzer werden ihre Daten nicht mehr nur von einem Standort aus abrufen, sondern werden sich über verschiedene Zugangspunkte in ihr Netz einwählen wollen. Die Benutzer können sich jedoch nicht immer darauf verlassen, dass der Anbieter eines mobilen Zugangs ihnen eine sichere Datenübertragung zu Verfügung stellt. Neben Antivirus Applikationen und Personal Firewalls müssen sich die Benutzer nun auch mit netzwerktechnischer Sicherheit auseinandersetzen.

6.4.1 Verschlüsselungen und weitere Verfahren

Viele der unten genannten Sicherheitsmassnahmen, die für mobile Datenkommunikation einsetzbar sind, gelten auch in anderen Bereichen der Informatik. Zudem kann Sicherheit verschieden definiert werden: Sicherheit ist der Zustand des Nichtvorhanden- oder geschützt seins vor Bedrohung und Risiken oder Sicherheit ist immer situationsbezogen

(d.h. abhängig von der aktuellen Bedrohungslage) [13]. So ist die Authentifikation und Autorisierung auch überall sonst anwendbar und es wird hier nicht weiter darauf eingegangen. Interessanter sind Wireless, spezifische Sicherheitsmassnahmen und Verschlüsselungen, welche eine sichere Datenübertragung gewährleisten können.

WEP (Wired Equivalent Privacy) Verschlüsselung

WEP ist der Standardverschlüsselungsalgorithmus für WLANs. WEP soll sowohl den Zugang zum Netz regeln, als auch die Geheimhaltung und Integrität der Daten sicherstellen.

Funktionsweise

Die WEP-Verschlüsselung verwendet einen gemeinsamen geheimen Schlüssel und den RC4 Verschlüsselungsalgorithmus. Der Access Point und alle mit ihm verbunden Stationen müssen denselben gemeinsamen Schlüssel verwenden. Für jedes gesendete Paket kombiniert der Sender die Inhalte des Paketes mit einer Checksumme des Paketes. Der WEP-Standard fordert den Sender dann auf, einen paketspezifischen Initialisierungsvektor zu erstellen. Dieser wird mit dem Schlüssel kombiniert und für die Verschlüsselung des Paketes verwendet. Der Empfänger erstellt seinen eigenen passenden Paketschlüssel und verwendet diesen, um das Paket zu entschlüsseln. In folgender Abbildung 6.6 wird dies weiter verdeutlicht.

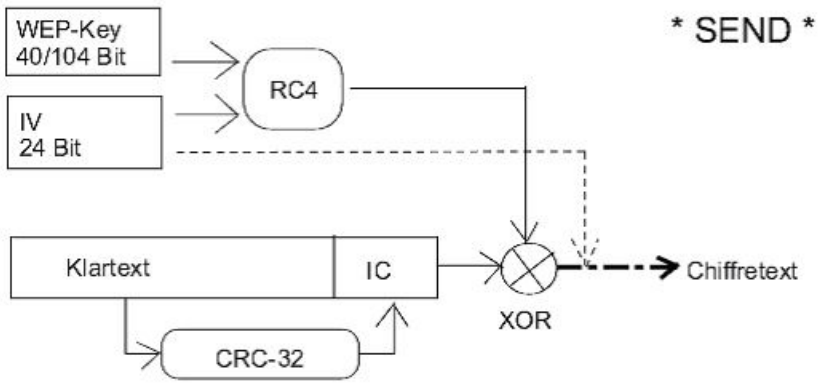

Abbildung 6.6: Funktionsweise WEP Verschlüsselung

Verwundbarkeit von WEP

WEP hat mehrere Schwachstellen, angefangen von der Verwendung gemeinsamer Schlüssel aller Stationen im Netzwerk. Wird der Schlüssel geknackt, so kann der Angreifer den Netzverkehr belauschen und erhält auch Zugang zum Netzwerk. Die Länge des WEP Schlüssel sowie auch die Länge des Initialisierungsvektors sind zu kurz und können mittels brute Force Methode geknackt werden. Die WEP Verschlüsselung erlaubt zudem auch die Wiederverwendung des Initialisierungsvektors. Durch die zu kurze Länge des Initialisierungsvektors (24Bit) ist, wenn der generierte Bitstrom für zwei Datenpakete unterschiedlich ist,

nach 2hoch24 = 16.8 Mio. verschiedenen Schlüsseln die erste Wiederholung des Initialisierungsvektors zu erwarten, vorausgesetzt, der IV wird zufällig initialisiert. Dadurch wird ein Angriff auf WEP vereinfacht, weil ein Angreifer nach einer gewissen Zeitspanne die gleiche Verschlüsselung knacken muss. In der Praxis bietet WEP keine Möglichkeit den Schüssel automatisch zu tauschen. Die Einstellung der Schlüssel muss manuell an allen Stationen erfolgen, das führt dazu, dass WEP Schlüssel selten bis nie geändert werden. [14] [16]

Tools

Im Internet sind zudem schon Tools erhältlich, mit welchen WEP Verschlüsselungen geknackt werden können. Ein Beispiel eines Tools das verwendet werden kann, um die WEP Verschlüsselung zu knacken ist AirSnort. AirSnort horcht passiv den WLAN Datenverkehr ab. Dieses Toll benötigt nach Angaben der Vertreiber zwischen 5 und 10 Millionen verschlüsselter Datenpakete. Wurden diese durch den mitgehorchten Datenverkehr gesammelt, so soll das Programm den WEP Schlüssel unter einer Sekunde erraten können. [27] [28]

Fazit

Ein Angreifer sucht immer nach der schwächsten Stelle, um einen Angriff zu starten. Bei eingeschalteter WEP Verschlüsselung ist der Aufwand, der ein potentieller Angreifer betreiben muss, um in ein Netzwerk einzubrechen, wesentlich grösser und somit wird auch die Wahrscheinlichkeit eines Angriffs reduziert. Obwohl, wie oben beschrieben, Tools bereitgestellt werden, die WEP Verschlüsselungen knacken können, muss bei eingeschalteter WEP Verschlüsselung immer noch Zeit investiert werden. So wird ein Angreifer vielleicht eher einen unverschlüsselten Zugang suchen, um seinen Angrifsversuch nicht unnötig komplizierter zu machen. Ökonomisch gesehen bewirkt die WEP Verschlüsselung für einen Angreifer einen Trade-off zwischen Suchzeit eines unverschlüsselten Zugangs und Zeit die er zum Datenpakete sammeln braucht und die Verschlüsselung eines Zuganges knacken zu können.

VPN (Virtual Private Network)

In einem Virtual Private Netzwerk wird eine auf TCP/IP-basierende Verbindung aufgebaut, welche für den Datenaustausch ein öffentliches Netz, wie das Internet benutzt. Die Verbindung über das öffentliche Netz wird durch Verschlüsselung gesichert. Die Verbindung wird als sog. Tunnel bezeichnet, welcher durch einen VPN-Client und einen VPN-Server aufgebaut wird. Der so aufgebaute Tunnel ist meist gesichert, es existieren jedoch auch Klartexttunnel.

Anwendung von VPN

VPN wirt oft dazu verwendet, um Mitarbeiter ausserhalb der Firma mit dem internen Netzt zu verbinden, damit sie so arbeiten können, als wären sie in der Firma selber. Diese Massnahme wird auch verwendet um WLANs und andere Funkstrecken zu sichern und wird als End-to-Site VPN bezeichnet.

Werden zwei lokale Netzwerke miteinander verbunden, wie es bei Firmen mit mehreren Geschäftsstellen der Fall sein kann, so wird auf beiden Seiten ein VPN-Gateway verwendet. Diese Verwendung von VPN nennt sich Site-to-Site VPN.

Wird ein Tunnel zwischen zwei einzelnen Computern aufgebaut, was in der Praxis selten der Fall ist, so wird dies als End-to-End VPN bezeichnet. [18] [17]

Sicherheit

Durch die Verwendung geheimer Schlüssel, Zertifikaten und Passwörtern kann die Authentizität sichergestellt werden. Eine durch einen VPN-Tunnel gesicherte Verbindung garantiert jedoch keinerlei Sicherheit vor internen Angriffen. Gegen Missbräuche welche, schon durch auf dem Rechner befindlichen Trojaner verursacht werden, bietet VPN keine Sicherheit.

6.4.2 Starke Passwörter

Die Sicherheit in der ganzen Informatik und damit auch beim mobilen Datenverkehr hängt stark von Passwörtern ab. Es wird jedoch immer wieder aufgezeigt, dass Benutzer leichtfertig immer gleiche oder leicht knackbare Passwörter verwenden. Der Umgang mit Passwörter wird für die Benutzer immer umständlicher. Ein durchschnittlicher Benutzer benötigt schon heute zahlreiche Passwörter, da scheint es auch nicht verwunderlich, dass die Sicherheit vernachlässigt wird.

Die Datenschutzbeauftragten des Kantons Zürichs[19] stellen zurzeit auf ihrer Homepage ein Tool bereit, mit dem Benutzer die Stärke ihres Passwortes testen können. In ihrer Auswertung, die über die letzten 12 Monaten reicht, wurden über 360'000 Passwörter überprüft. Die Ergebnisse sind erschreckend, die Mehrzahl der geprüften Passwörter wird als schwach und somit unbrauchbar eingestuft.

Wieso sind besonders bei mobiler Kommunikation starke Passwörter wichtig? Bei Funkübertragung, anders als bei der Datenübertragung durch Kabel, werden die Daten radial vom Funksender ausgesandt und müssen von einem Empfänger empfangen werden. So kann theoretisch jeder mit einem geeigneten Empfänger den Datenstrom abhören. Starke Passwörter stellen nun sicher, dass ein Angriff durch den grösseren Aufwand verlängert wird. Doch was ist ein starkes Passwort? Ein Passwort sollte sicher nur einer Person bekannt sein und auch nur von ihr verwendet werden. Der Benutzer sollte es sich einfach merken können und ein potentieller Angreifer soll das Passwort nicht erraten dürfen. Das

Passwort sollte aus mehr als 8 Zeichen bestehen, Gross- und Kleinschreibung, sowie Zahlen und Sonderzeichen beinhalten. Man sollte sie auch leicht eintippen können, um so die Gefahr des Ausspähens zu minimieren.[20]

6.4.3 Sicherheit bei Mobiltelefonen

Mit der Viren- und Spamplage bei Computern haben wir uns mittlerweile schon abgefunden und nicht daran gedacht, dass auch Mobiltelefone ihre Schwachstellen besitzen.

Bluetooth

Bluetooth wird angepriesen für einfachen Datenaustausch zwischen verschiedenen Geräten, welche sich in unmittelbarer Distanz befinden. Bluetooth ist aber bei Mobiltelefonen eine unterschätzte Sicherheitslücke. Forscher der Salzburg Research Forschungsgesellschaft mbH haben es geschafft Mobiltelefone, besonders Nokia 6310 über eine Distanz von 1.7 Km zu manipulieren. Ihnen ist es gelungen, Nummern anzuwählen, SMS zu beliebigen Nummern zu schicken, den SMS Speicher auszulesen, Telefonbucheinträge zu lesen und zu erstellen. Ihnen ist es sogar gelungen über diese Distanz eine Verbindung zum Internet aufbauen, welche es ihnen unter anderem ermöglicht hätte Viren und Würmer zu verbreiten.[24] [25]

Ein wichtiger Schritt zu einer Verbesserung der Sicherheit bei Mobiltelefonen ist es, die Sichtbarkeit (Discovery Mode) für die Gegenstelle abzuschalten. Geräte, wie zum Beispiel eine Freisprechanlage, welche schon mit dem Mobiltelefon gekoppelt sind, brauchen diese Sichtbarkeit nicht mehr. Eigene Versuche, mit der Suchfunktion des Mobiltelefons in grossen Menschenansammlungen, haben ergeben, dass viele Besitzer von Mobiltelefonen die Sichtbarkeit ihres Gerätes eingeschaltet lassen und damit zu einem potentiellen Angriffsziel werden. [21]

Java Virtual Maschine

Viele der Betriebssysteme der heutigen Mobiltelefone sind multitaskingfähig und erlauben es, Anwendungen als ständige Hintergrundprozesse zu betreiben. Durch die grosse Anzahl von Kommunikationskanälen wie zum Beispiel TCP/IP per GSM, UMTS, Bluetooth und neuerdings auch noch WLAN ist die Angriffsfläche von Mobiltelefonen erheblich grösser, als die bei den Personalcomputern.

Die Java Virtual Maschine sind Sicherheitsmassnahmen eingebaut, welche den Missbrauch verhindern sollen. In der KMV (Kilobyte Virtual Maschine), eine JVM, speziell entwickelt für PDA's und Mobiltelefone, verzichtete SUN aus Performancegründen auf eine Sicherheitskontrolle des vom Java-Compilererzeugten Bytecodes zur Laufzeit. Laut dem Bericht von Heise Security bestätigte ein Sun Sprecher, dass diese Lücke bekannt sei. Sie hätten die bestehende Lücke bereits geschlossen und allen Lizenznehmern eine neue Version zur

Verfügung gestellt. Durch die Fehlüberlegung, bei der Konzeption von KMV, ist es möglich mit Hilfe von Trojaner getarnt als Spiele oder anderen Applikationen, wie das unten genannte Beispiel zeigen wird, Persönliche Daten zu kopieren, zu löschen und Einträge zu generieren. Szenarien werden aufgezeigt, in denen teure Internetverbindungen aufgebaut werden. Ein Trojaner wäre sogar in der Lage, den internen Flashspeicher zu überschreiben, um so das Mobiltelefon unbrauchbar zu machen.[22] [23]

Beispiel eines Handy Trojaners

Das Magazin iX hat als Proof-of-Concept eine Spyware für Mobiltelefone geschrieben. Eine kleine Applikation, getarnt als ein Teatimer. Beim Starten der Applikation kopiert sie alle Einträge des Adressbuches in eine Textdatei. Wechselt der Benutzer nun in ein anderes Programm, so wird der Trojaner nicht beendet, sondern verbleibt im Hintergrund und überwacht weiter den SMS Eingang. Empfängt das infizierte Mobiltelefon eine SMS mit einem bestimmten Präfix und einer gültigen Mobilnummer, so schickt es die erstellte Textdatei an die gegebene Nummer und löscht die erhaltene SMS aus dem Eingangsordner.

Lösungen

Die ersten Lösungen zur Bekämpfung des, immer mehr aufkommenden, Sicherheitsproblems von Mobiltelefonen scheinen sich zu entwickeln. Symantec will sich bemühen ihre Sicherheitskonzepte auch in mobilen Kleingeräten unterzubringen. Soll es nach Symantec gehen, müssen Benutzer in Zukunft ihre Mobiltelefone in gleicher Weise schützen, wie man es heutzutage mit den Personalcomputern gewohnt ist. Die Firma Symantec hat in Zusammenarbeit mit der Firma Nokia eine Sicherheitslösung für den Nokia Communicator 9500 ausgearbeitet. Die Sicherheitslösung beinhaltet eine Firewall, um ein- und ausgehende Datenpakete überprüfen und evt. blockieren zu können. Einen Echtzeit Virenschutz, welcher mittels WLAN, durch sogenannte LiveUpdates, immer auf dem neusten Stand gehalten wird. Durch diese Sicherheitslösung soll es auch möglich sein, die Sicherheitsrichtlinien eines Unternehmens auch auf die mobilen Geräte ausweiten zu können, um so ein kompaktes Sicherheitskonzept zu bewirtschaften. [26] [29] [30]

6.4.4 Zertifikate

Zertifikate ist ein Begriff der nicht nur in der Informatik beheimatet ist. Zertifikate werden benutzt um die Integrität und Echtheit von Inhalten zu überprüfen. Die Anwendung von Zertifikaten ist schon weit verbreitet, doch wird sich die Ausbreitung noch weiter ausdehnen. Bei Zertifikaten ist es ausgesprochen wichtig, dass eine zentrale Stelle vorhanden ist, der alle Zertifikatbenutzer vertrauen können. Diese Organisation kann durch eine private Firma oder auch durch eine staatliche Behörde betrieben werden.

Beispiel eines Zertifikates

Das folgende Beispiel ist ein X-509 v3 Zertifikat. Die Strucktur eines Solchen Zertifikats ist wie folgt aufgebaut:

- Version: Hier zum Beispiel Version 3
- Seriennummer
- Algorithmen ID
- Aussteller:
- Gültigkeit
- Von
- Bis
- Subject
- Subject Public Key Info
- Public Key Algorithmus
- Subject Public Key
- Eindeutige ID des Austellers (optional)
- Eindeutige ID des Inhabers 8optional)
- Erweiterungen
- Zertifikat Signaturalgorithmus
- Zertifikat Signatur

Mithilfe der Struktur des X-509 v3 ist es leicht das Beispiel (Abbildung 6.7)eines Zertifikates zu verstehen [31].

6.5 Disruptive Technologies

Der Begriff disruptive Technology wurde zum ersten Mal durch den BWL-Professor Clayton M. Christensen[33] der Harvard Business School, in dem Buch The Innovators Dilemma benutzt. Disruptive Technologien können zeigen, ob ein Geschäftsmodell oder eine neue Erfindung, das Potential beinhaltet sich im Markt zu etablieren und so den gängigen Technologien die Führung in Frage zu stellen.

```
Certificate:
  Data:
    Version: 3 (0x2)
    Serial Number: 1 (0x1)
    Signature Algorithm: md5WithRSAEncryption
    Issuer: C=XY, ST=Austria, L=Graz, O=TrustMe Ltd, OU=Certificate Authority, CN=CA/Email=ca@trustme.dom
    Validity
      Not Before: Oct 29 17:39:10 2000 GMT
      Not After : Oct 29 17:39:10 2001 GMT
    Subject: C=DE, ST=Austria, L=Vienna, O=Home, OU=Web Lab, CN=anywhere.com/Email=xyz@anywhere.com
    Subject Public Key Info:
      Public Key Algorithm: rsaEncryption
      RSA Public Key: (1024 bit)
        Modulus (1024 bit):
          00:c4:40:4c:6e:14:1b:61:36:84:24:b2:61:c0:b5:
          d7:e4:7a:a5:4b:94:ef:d9:5e:43:7f:c1:64:80:fd:
          9f:50:41:6b:70:73:80:48:90:f3:58:bf:f0:4c:b9:
          90:32:81:59:18:16:3f:19:f4:5f:11:68:36:85:f6:
          1c:a9:af:fa:a9:a8:7b:44:85:79:b5:f1:20:d3:25:
          7d:1c:de:68:15:0c:b6:bc:59:46:0a:d8:99:4e:07:
          50:0a:5d:83:61:d4:db:c9:7d:c3:2e:eb:0a:8f:62:
          8f:7e:00:e1:37:67:3f:36:d5:04:38:44:44:77:e9:
          f0:b4:95:f5:f9:34:9f:f8:43
        Exponent: 65537 (0x10001)
    X509v3 extensions:
      X509v3 Subject Alternative Name:
        email:xyz@anywhere.com
      Netscape Comment:
        mod_ssl generated test server certificate
      Netscape Cert Type:
        SSL Server
  Signature Algorithm: md5WithRSAEncryption
    12:ed:f7:b3:5e:a0:93:3f:a0:1d:60:cb:47:19:7d:15:59:9b:
    3b:2c:a8:a3:6a:03:43:d0:85:d3:86:86:2f:e3:aa:79:39:e7:
    82:20:ed:f4:11:85:a3:41:5e:5c:8d:36:a2:71:b6:6a:08:f9:
    cc:1e:da:c4:78:05:75:8f:9b:10:f0:15:f0:9e:67:a0:4e:a1:
    4d:3f:16:4c:9b:19:56:6a:f2:af:89:54:52:4a:06:34:42:0d:
    d5:40:25:6b:b0:c0:a2:03:18:cd:d1:07:20:b6:e5:c5:1e:21:
    44:e7:c5:09:d2:d5:94:9d:6c:13:07:2f:3b:7c:4c:64:90:bf:
    ff:8e
```

Abbildung 6.7: Beispiel eines Zertifikates X-509 v3

Eine disruptive Technologie ist ein Produkt oder ein Prozess, mit einer niedrigeren Performance oder einem niedrigerem Preis, als bereits bestehende Produkte. Diese neue Technologie beginnt am unteren Ende eines existierenden Marktes, wo wenig Druck und Konkurrenz existiert. Durch Performancesteigerung und Preisnachlass steigt es etappenweise den Markt hoch und ersetzt am Ende das marktdominierende Produkt. Normalerweise werden diese disruptiven Technologien durch startup Unternehmen oder Neueinsteiger in den Markt eingebracht.[34]

Um herauszufinden ob eine Technologie zu der Kategorie der disruptiven Technologien gehört, muss man die Kriterien aufstellen, die eine Technologie als disruptiv Kategorisieren. Es sind keine eindeutigen Kriterien, da der Begriff „disruptive Technologie“ immer noch nicht genau definiert wurde. Die hier aufgeführten Kriterien sind aber allgemein akzeptiert und werden in ihrer Allgemeinheit benutzt um neue Technologien als disruptiv einzustufen.

- Neue Technologie
- Billigere Technologie
- Weniger Performance als bestehende Technologien

- Wächst die Performance stetig? Wird die Performancesteigerung andere Produkte übertreffen?
- Ersetzt sie eine andere Technologie?

6.5.1 Die Entstehung neuer Technologien

In einer sich schnell entwickelnden technologischen Gemeinschaft, werden neue Technologien so rasant entwickelt, dass sich die Produktneuerungen schwer durch Benutzer integrieren und verwerten lassen. Man kann sagen, dass diese Neuerungen die Bedürfnisse der Benutzer übersteigen und auch übertreffen. Die neuen Technologien werden nicht akzeptiert oder gar nicht bemerkt. Zu diesem Zeitpunkt, steigt diese neue Technologie, mit einer niedrigeren Performance als die bestehenden Technologien, in den Markt ein, um schnell an Kunden und an Markt zu gewinnen. Um sich nicht in einen Preiskampf mit den neuen Technologien einzulassen, fokussieren die bestehenden Unternehmen auf spezielle Marktnischen, in der die Benutzer Wert auf grössere Performance legen. Nach einigen Runden, wird der alte Marktinhaber in sukzessiv kleinere Marktsegmente gedrängt, bis die alte Technologie letztendlich verschwindet.

Aber nicht alle disruptiven Technologien haben eine niedrigere Performance als ihre Vorgänger. Es existieren mehrere verschiedene Beispiele, in denen die neue Technologie nicht von den marktdominierenden Unternehmen akzeptiert und übernommen wird. Viele dieser Unternehmen haben grosse Summen in die alten Technologien investiert und müssten durch einen Wechsel die Infrastruktur ersetzen oder sogar, diese mit hohen Kosten entfernen lassen. Meistens ist es von grösserem Vorteil, die Gewinne der sterbenden Technologie bis zum letzen auszunutzen, als sie durch neue Investitionen zu verbessern. Neueintreter haben keine solchen Probleme und können direkt in die neue Technologie investieren.

Ein anderes Beispiel sind Unternehmen die früher selbst mal Innovatoren waren und in der ersten Generation erfolgreich die etablierten displazierten. Diese Unternehmen setzen sich mit ihrem eigenen Erfolg fest, so dass sie den Wechsel in die neuen Generationen verpassen. So wie sie die ehemaligen Marktführer disruptiert haben, werden sie durch die neu erscheinenden Technologien der nächsten Innovationswelle disruptiert.

6.5.2 Beispiele disruptiver Technologien

Ein oft genanntes Beispiel einer disruptiven Technologie ist das Telefon. Im 19. Jahrhundert dominierte Western Union den Telekommunikationsmarkt mit der Kommunikation über Telegraphen. Als Bell mit dem Telefongeschäft begann, entsprach die Sendeleistung des Telefons nur 5 Km, aber Telegraphen konnten über hunderte von Kilometer kommunizieren. So konnte sich das Telefon nur in kleinen Nischen behaupten. Das grosse Geschäft blieb aber Western Union vorbehalten. Als die Technologie des Telefons einen Sprung machte und auch über grosse Distanzen hinweg gute Qualität behielt, konnte Bell, Western Union vom Kommunikationsmarkt verdrängen und so die Telegraphen dem Untergang widmen[35].

Ein anderes Beispiel, in dem der Wandel noch in der Mitte steht, ist die Substitution der Filmfotographie durch die digitale Fotographie. Die digitale Fotografie hat klein Angefangen. Die ersten Kameras hatten eine miserable Auflösung, so dass dies am Anfang dazu führte, dass nur Technologie-Enthusiasten diese Kameras kauften. Aber als die Qualität und die Performance der Kameras weiterhin stieg und auch der Preis attraktiver wurde, entwickelte sich die digitale Kamera zu einem signifikanten Konkurrenten für die bestehende Technologie. Heute kann man sagen, dass die digitale Fotografie die Filmfotographie verdrängt hat, obwohl die Qualität der digitalen Bilder immer noch nicht an die der Filme angelangt ist. Ein gutes Beispiel für eine Firma die den Wandel der Zeit bemerkt hat ist Kodak, die ihre ganze Produktion für Film-Fotographie stilllegen liess und auf digitale Fotografie umgestiegen ist.

Weitere Beispiele disruptiver Technologien:

- Buchdruck vs. Manuskripte
- ADSL vs. ISDN
- Autos vs. Pferdekutschen
- IP vs. Andere proprietäre Netzwerkprotokolle

Es gibt auch einige Beispiele, wo der Prozess, die ältere Technologie zu verdrängen, noch im vollen Gange ist und man nicht mit Gewissheit sagen kann ob sie wirklich verschwinden wird oder nicht.

- VoIP vs. Analoge oder digitale Festnetze
- Mobile Telefonie vs. Festnetztelefonie
- Download von Musik (MP3) vs. CD´s
- E-Book vs. Papierbücher
- Open Source Software vs. proprietäre Software

Es gibt auch Beispiele von Technologien die ursprünglich als disruptiv bezeichnet wurden, aber letztendlich von anderen Technologien überholt oder nicht durch die Benutzer akzeptiert wurden.

- Betamax
- Laserdisc
- Virtual Reality

6.5.3 WLAN: eine disruptive Technologie?

Um herauszufinden ob WLAN zu der Kategorie der disruptiven Technologien gehört, werden die Eigenschaften des WLANs mit den vorher aufgestellten Kriterien verglichen und evaluiert.

Man kann sagen, dass das WLAN relativ neu ist und seine Marktpenetration erst ab Ende 2002 begonnen hat, obwohl der 802.11-Standard schon seit 1997 existiert. Erst ab der Implementierung des 802.11b Standards fand WLAN seinen Durchbruch.

Die Frage ob WLAN billiger ist kann man nicht genau beantworten, da in diesem Fall ein exaktes Gegenstück fehlt. Wenn man bedenkt, dass man für die Installation von Ethernet-Lan meterlange Kabel benötigt und manchmal auch Wände und Mauern durchbohren muss, kann man sagen, dass WLAN billiger ist. Wenn man Ethernet-Verbindungen mit WLAN vergleicht, muss man auch feststellen das letzteres weniger Performance aufweist. Im Gegensatz zu 100-Megabit oder Gigabit-Lan, weist der letzte 802.11-Standard maximal 55 MBit/s Transfergeschwindigkeit auf.

Die Performance wächst auch stetig. Seit dem ersten Standard, dass maximal 1 MBit/s ermöglichte, ist der Standard bis auf 55 MBit/s gestiegen. Der nächste Standard (802.11n) soll bis Ende 2006 verabschiedet sein und Datenraten bis über 100 MBit/s ermöglichen und auch grössere Distanzen bieten. Zu diesem Zeitpunkt wird WLAN auch die weit verbreiteten 100Mbit Netzwerke überholen.

Mit dieser stetigen Performancesteigerung wird WLAN bald kabelgebundene Netzwerke ersetzen. Heutzutage ergänzen sich beide Netzwerkstandards noch, aber in Zukunft wird die Entscheidung für WLAN immer öfter fallen. Vor allem in Haushalten und kleinen Firmen, die bis jetzt noch nie an die Installation eines Netzwerkes gedacht haben, entfacht WLAN neue Möglichkeiten verschiedene Computer miteinander zu verbinden. WLAN bietet auch eine grössere Freizugigkeit gegenüber kabelgebundenen Netzwerken, da man nicht an einen spezifischen Ort gebunden ist um in das Netzwerk zu gelangen.

Wenn man alle Punkte vereinigt und berücksichtigt kann man mit Gewissheit sagen, dass WLAN in Haushalten und KMU's zu einer disruptiven Technologie wird, währenddessen in grossen Unternehmen, wo der Bedarf an grosser Bandbreite herrscht, WLAN eher als komplementäre Technologie angesehen wird.

6.6 Verteidigungsstrategien der Etablierten

Die Verteidigungsstrategien der Etablierten sind Massnahmen der Unternehmen, die marktdominierend sind, um Neueintreter oder Substitutoren von ihrem Markt zu verdrängen. Sie zielen darauf neue disruptive Technologien zu schnellstmöglich zu erkennen, und wenn diese eine Gefahr bilden, sie möglichst effektiv zu bekämpfen.

In jedem Wirtschaftsbereich, sei es im Inland oder international, müssen sich die Unternehmen dem wachsenden Wettbewerb stellen. Michael Porter[36] hat in seinem Buch

`Competitive Strategy`, vor 20 Jahren schon die 5 Kräfte aufgezeichnet, denen ein Unternehmen im Wettbewerb ausgesetzt ist.

- *Der Markteintritt durch neue Wettbewerber:* Neue Technologien und Wettbewerber machen entsprechende Reaktionen der etablierten Unternehmen notwendig um sich gegen diese neuen Produkte zu wehren. Dabei sind Investitionen in neue Ressourcen und Gewinnschmälerung unausweichlich.
- *Die Herausforderung durch Ersatzprodukte oder substitutiven Technologien:* Wenn es auf dem Markt gängige Alternativen zum Produkt oder Technologie einer Unternehmung gibt, die die volle Ausschöpfung der Gewinne begrenzt.
- *Die Verhandlungsmacht der Käufer:* Kunden, die in den Wettbewerb einsteigen oder über grosse Verhandlungsmacht verfügen könne Gewinne der etablierten Unternehmen schmälern.
- *Die Verhandlungsmacht der Verkäufer:* Haben die Lieferanten eine grosse Verhandlungsmacht so können sie diese ausnutzen um Teile vom Gewinn der etablierten Unternehmen zu stehlen.
- *Der interne Wettstreit im Markt:* Der interne Wettbewerb erfordert massive Investitionen in Forschung und Entwicklung sowie ein agressives Marketing von Seite der Unternehmen.

Die Vereinigung dieser 5 Mächte des Wettbewerbs ermöglicht es einem Unternehmen sich im Wettbewerb zu behaupten und zu überleben. Um dies zu bewerkstelligen, müssen die Unternehmen Markteintrittsbarrieren schaffen, die es den neuen Unternehmen unmöglich machen, sich im Markt zu etablieren und dadurch Gewinne zu kassieren.

6.6.1 Das Individuum als Kraft

Durch die Entschlossenheit der Individuen eine Technologie zu akzeptieren und sie zu benutzen, entwickeln diese, ohne es zu wissen, eine Macht gegenüber gewissen Unternehmen und Technologien. Ist diese Technologie auch noch kostengünstig und frei verfügbar, stellt sie eine ernsthafte Bedrohung den etablierten Unternehmen dar. Das Individuum ist Kunde und Substitutor sogleich, wie in Abbildung 6.8 veranschaulicht.

Im Falle des WLANs geschieht genau dies. Der WLAN Standard ist für alle frei zugänglich und die Benutzung ist, bis auf wenige Ausnahmen, kostenlos. Das Individuum das früher, nur Kunde, der etablierten Unternehmen war, nutzt eine neue substitutive Technologie um Geld zu sparen und erhält dadurch die Macht mit diesen Unternehmen zu verhandeln und auf die Preise zu drücken.

Treten in diesem Fall noch Netzwerkexternalitäten auf, d.h., der Nutzen für jeden neuen Konsumenten einer Technologie steigt, je mehr Personen diese Technologie benutzen, so können die Unternehmen nicht anderes machen, als Verteidigungsstrategien aufzubauen, um sich vor solchen Technologien zu schützen.

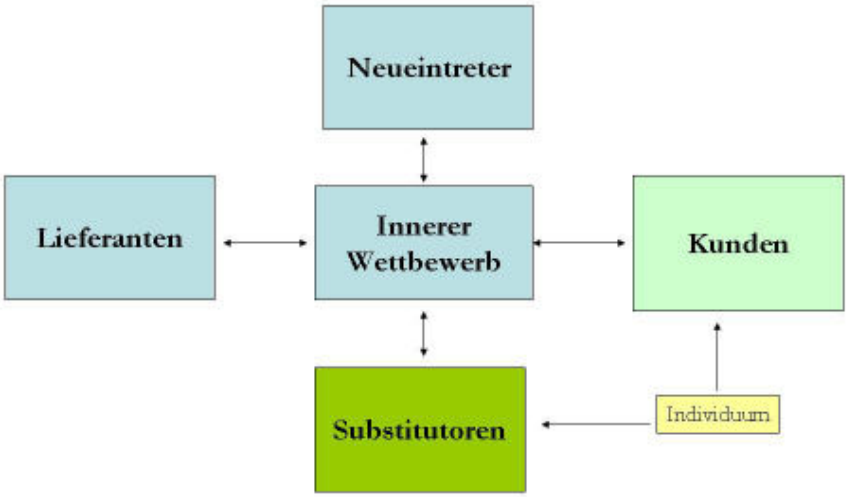

Abbildung 6.8: Die 5 Kräfte von Porter im Gegensatz zum Individuum

6.6.2 Markteintrittsbarrieren

In den vergangenen 10 Jahren haben die Telekommunikationsunternehmen viel in neue Technologien investiert. Darunter kann man GSM, GPRS und in naher Vergangenheit UMTS nennen. Diese Investitionen müssen erst amortisiert werden und Gewinne erwirtschaften. Vor allem die Investitionen in Lizenzen für UMTS, haben die Reserven an liquiden Mitteln schrumpfen lassen. Die grossen Unternehmen verhalten sich im Moment sehr vorsichtigt. Grössere Investitionen werden im Moment nicht getätigt. Zu klein ist noch die Bedrohung, die von anderen, offenen Technologien hervorgeht. Zu diesem Zeitpunkt ist die beste Strategie, die die Marktführer verfolgen können, der Ausbau ihrer Kapazitäten um ihre Marktposition zu festigen und möglicherweise auch auszubauen. Desto stärker die Unternehmen in ihren Bereichen sind, umso schwieriger wird es, für Neueinsteiger oder Substitutoren einen Fuss im Markt zu fassen. Um ihre Position im Markt auszubauen, müssen die etablierten Unternehmen Services fördern welche die Netzwerkexternalitäten eines Marktführers aufbauen und expandieren. Solche Services können verschiedene Charakteristiken haben, müssen aber den Dienstleistungscharakter des Unternehmens nicht gefährden und mit ihrem Image übereinstimmen. Speziell in der Telekommunikationsbranche, gibt es immense Möglichkeiten solche Services anzubieten. Nicht nur für Mobilnetze sondern auch für alle Technologien die mit der Mobilfunktechnik eng zusammenarbeiten. Mit dem Zusammenschluss der Technologien, sei es Mobilfunk mit Bluetooth, Internet mit WLAN, Mobilfunk mit Internet usw. ist es möglich verschiedene Angebote dem Kunden anzubieten, die die Integration dieser Technologien beinhalten.

6.6.3 Beispiele von Services, die die Netzwerkexternalität erhöhen

1. Ein Beispiel von Service, wäre das Angebot von speziellen Programmen, die nur bei einem Anbieter verfügbar sind und nur über das eigene Netzwerk der etablierten Unternehmen zugänglich sind. Z.B. ein Online-Spiel, das nur über das Netzwerk von Swisscom gespielt werden kann.

2. Ein weiteres Beispiel könnte die Blockierung von fremden Services sein, um eigene mServices zu fördern. Z.B. die Blockierung der Bluetoothübertragung zwischen Computer und Handy bei einigen mobilen Apparaten der Marke Vodafone. Nur die Übertragung zwischen speziellen Freisprecheinrichtungen über Bluetooth ist gestattet. Somit fördert Vodafone die Benutzung der eigenen Internet-Services, die gegen eine hohe Gebühr entrichtet werden müssen. Z.B. das Übertragen von Musik auf das Handy, gelingt nicht per Bluetooth, man muss die Musik von Vodafone direkt oder über ein spezielles Extrakabel übertragen, was man natürlich nicht gratis mitgeliefert bekommt. In weiterer Zukunft ist es auch vorstellbar, dass die neue Generation der WLAN fähigen Handys auch für fremde Netze blockiert werden könnten. So könnte z.B. eine Person die bei Swisscom Kunde ist, nur über die Hotspots der Swisscom oder ihren Partnern telefonieren. Würde man versuchen über andere WLAN Netze zu telefonieren, könnte man blockiert oder für Roaming zur Kasse gebeten werden.

3. Ein weiteres Beispiel von Service, ist die verbilligte Benutzung von Hotspots eines Unternehmens, wenn man schon einen Internetvertrag oder Anschluss bei jeweiligem Unternehmen besitzt. Dies ist heute schon der Fall bei der Firma Cablecom, die den Kunden ihres Breitbandangebots, eine verbilligte Benutzung ihres Angebots an Hotspots offeriert. Zwar ist der Preis immer noch relativ hoch, aber mit der weiteren Konkurrenz in diesem Markt, steht dem Preisverfall für die Accesspointbenutzung nichts im Wege.

4. Ein häufiger Fall von Förderung von Netzwerkexternalität ist die Bereitstellung oder Schenkung von Hardware, mit dem Abschluss eines Vertrages bei einem Unternehmen. Bei Telefongesellschaften ist dies die verbilligte Abgabe von Handys an vertragsabschliessende Kunden. Bei Breitbandanbietern, die Schenkung von Modems und Routern.

6.6.4 Investition in neue Technologien

Um im Wettbewerb bestehen zu können müssen die etablierten Firmen ständig in neue Technologien investieren, um eine grössere Vielfältigkeit an Services den Kunden anbieten zu können. Natürlich versprechen sich diese Unternehmen einen grösseren Umsatz durch diese neuen Angebote. Die Investition in Infrastruktur und Technologie erhöht nicht nur den Umsatz, sondern schützt sie auch vor neuen Technologien, die als Ersatz für die bestehen Produkte in den Markt eindringen könnten.

Eine dieser neuen Technologien ist das Universal Mobile Telecommunications System, besser bekannt als UMTS. UMTS wird auch als die dritte Generation (3G) des Mobil-

funks bezeichnet und basiert auf dem WCDMA-Verfahren (Wideband Division Multiple Access). Durch UMTS versprechen sich die Telekommunikationsunternehmen einen neuen Sprung an Qualität und Quantität an Services, die sie ihren Kunden offerieren können. Eine weitere Technologie, die in naher Zukunft von den Telekommunikations-Unternehmen ins Visier genommen werden könnte ist WiMax (Worldwide Interoperability for Microwave Access). WiMax ist ein neuer Standard (IEEE 802.16a) für lokale Funknetze, und verspricht bessere Übertragungsraten und ist über grössere Distanzen erreichbar als WLAN. Er wird aber heutzutage noch nicht kommerziell eingesetzt und ist nur in weiteren Expertenkreisen bekannt.

Vor- und Nachteile der Investitionen in UMTS

Seit der umstrittenen Lizenzvergabe in vielen europäischen Ländern, ist UMTS ein Wort, das bei vielen Experten gefürchtet wird. UMTS wurde früher als ein Wunder der Technik propagiert und würde die Telekommunikation auf ein neues Niveau katapultieren. Seit dem Rückgang des Internetbooms sieht die Sache anders aus. Durch die hohen Investitionen in Lizenzen und dem teueren Aufbau des Netzes, sind die Kommunikationsunternehmen zunehmend in Bedrängnis geraten, da jetzt auch die Technologie in Frage gestellt wurde. Langsam aber tritt Bewegung in den Markt ein. Einige europäische Länder, darunter auch die Schweiz, haben ihre UMTS-Netze bereits aufgeschaltet und versuchen die ersten Erfolge zu ernten. Darunter preisen die Anbieter verschiedene Vorteile von UMTS, im Vergleich zu anderen Technologien wie WLAN oder GPRS an. Ein erster Vorteil den man Nennen kann, ist die hohe Service-Qualität. Die Telekommunikationsunternehmen versuchen die Qualität auf dem Maximum zu halten und bemühen sich einen entsprechenden Kundensupport zu garantieren. Ein grosser Vorteil gegenüber WLAN und auch WiMax, ist der grosse Deckungsgrad den UMTS schon hat. In der Schweiz liegt der heute etwa bei 80% des Landes.

In diesem Zusammenhang ist auch das einwandfreie Roaming ein Pluspunkt gegenüber dem WLAN. Bewegt man sich bei einem WLAN Netz von einem Accesspoint zum nächsten, so muss man sich immer wieder neu in das Netzwerk einloggen. Auch die verschiedenen Services, die über UMTS anbietbar sind, stellen einen grossen Vorteil von UMTS dar. So ist es möglich Videokonferenzen, Online-Spiele oder Fernsehübertragungen auf dem mobilen Apparat zu empfangen. Ein weiterer Pluspunkt für UMTS, ergibt sich aus dem Thema Sicherheit. Während man bei privaten WLAN-Anbietern keine oder eine fast inexistente Sicherheit hat, kann man davon ausgehen, dass bei grossen Unternehmen die Sicherheit der Übertragung gewährleistet ist.

Als grossen Nachteil der UMTS-Technologie kann man heutzutage die hohen Preise für die angepriesenen Services nennen. Als Beispiel könnte man die 4 Franken pro halbe Stunde erwähnen, die die Swisscom für das mobile Fernsehen am Handy verlangt. Auch die geringere Bandbreite als WLAN oder WiMax wirkt sich negativ auf UMTS, da es im Moment auf maximal 2 Mbit/s übertragen kann. Diese Übertragungsrate fällt noch weiter ab, desto schneller sich die Person bewegt. Eine gute Bandbreite ist bei einer Autofahrt also fast unmöglich zu erreichen. Weiterhin ist zu erwähnen, dass eine Monopolsituation herrscht, da es nur wenige Anbieter auf dem Markt gibt, was im Gegensatz zu einem offenen WLAN-Netz liegt.

Vor- und Nachteile der Investitionen in WiMax

Obwohl WiMax noch im kommen ist, und fast keine komerziellen Netze existieren, kann man sagen, dass WiMax, wenn implementiert, eine grosse Konkurrenz zu WLAN darstellen könnte. WiMax übertrifft in vielen Aspekten WLAN. So hat es eine Reichweite von bis maximal 50 km (im Gegensatz zu 100m von WLAN). Aber es ist auch zu erwähnen, dass diese maximale Übertragungsdistanz nur theoretisch ist und in dicht besiedelten Gebieten bis zu 600 m abfällt[38].

Auch die Übertragungsrate ist in Theorie bis zu 70 Mbit/s schnell, fällt aber unter normalen Bedingungen bis zu 20MBit/s ab. Eigentlich sollte WiMax, ADSL und Kabelinternet ergänzen, z.B. in ländlichen Gebieten, wo beide Technologien unrentabel wären, aber WiMax wird auch des öfteren als Konkurrenz zu privat angebotenen WLAN-Netzen angesehen[37]. Ein Grund dafür, ist der hohe Investitionspreis, den nur Unternehmen aufbringen können um ein Netz zu installieren. Der Preis einer Basisstation ist heutzutage im gleichen Bereich wie die einer UMTS-Station. Eine Aussage von Michael Cai von Park Associate bringt die Sache auf den Punkt: *„Wi-Fi is really a consumer-driven business model, while WiMax is a service provider driven business model“.* [39] Ein weiter negativer Punkt für WiMax ist der noch nicht definitive Standard. So können Firmen die in WiMax investieren, nicht genaue Planungen anstellen. Würden sie in den falschen Standard investieren, könnte dies das aus dieser Firma bedeuten [40].

6.7 Zusammenfassung

In der heutigen technologischen Gesellschaft, in der alles miteinander integriert wird, und die Schnittstellen sich immer häufiger überlappen, ist es merkwürdig zu sehen, dass ein einzelnes Individuum zu einer Marktbedrohung werden kann. Durch sein Verhalten und durch seine Aktionen im technologischen Umfeld, bildet das Individuum ein Potential, durch einfache Mittel, grosse Operatoren zu gefährden. Durch die Verbindung einzelner Individuen zu einer Gemeinschaft, potenziert sich diese Gefahr. Durch diese Verbindungen entstehen variierte Geschäftsmodelle und Techniken, um Services dem Benutzer anzubieten, was vorher nur den grossen Marktunernehmen vorbehalten war. Ob diese Services in dieser Weise legal sind ist immer noch sehr unumstritten, aber einige rechtliche Grundlagen sind dabei schon zu beachten. Im Falle des WLANs, das in dieser Arbeit als Hauptbeispiel benutzt wurde, sind insbesondere das Fernmeldegesetz und auch das Strafgesetzbuch zu beachten. Diese beiden Gesetzesbücher regeln die Art und Weise wie die neuen Funktechnologien verbreitet und vermarktet werden müssen. Einen weiteren Aspekt den man beachten muss, ist die umfassende Sicherheit den ein Anbieter seinen Kunden offerieren muss. Heutzutage gibt es verschiedene Technologien für drahtlose Netzwerke, die die Sicherheit, nicht immer gewährleisten, aber wenigsten auf ein akzeptierbares Mass erhöhen können. Diese Sicherheit, offeriert von privaten oder öffentlichen Anbietern, kann sich aber nie mit der Qualität der grossen marktführenden Operatoren vergleichen. Diese setzen nämlich alle ihre Kräfte in die Verteidigung ihrer Marktführerschaft. Um dies zu bewerkstelligen investieren diese Unternehmen Milliarden in neue Technologien und Services um die Bemühungen der kleinen Anbieter schon im Keim zu ersticken. Beispiele

dafür sind die Investitionen in UMTS und WiMax. Durch diese Investitionen erhoffen sich die Unternehmen ihre Marktposition auszuweiten. Aber eine hundertprozentige Sicherheit werden sie nie erlangen, immer wieder treten neue disruptive Technologien auf, die mit einfachen Mitteln und technologischen Neuerung diese Verteidigungsstrategien der etablierten Unternehmen auf die Probe stellen werden.

6.8 Fazit

Um eine Gefährdung für die grossen Operatoren darzustellen, müssten die einzelnen Individuen besser miteinander kommunizieren. Die Verstreutheit der einzelnen Lösungen, sei dies von privaten oder öffentlichen Anbietern einzelner Services, bietet in diesem Moment noch keine Gefahr für die grossen Unternehmen. Auch rechtliche und sicherheitstechnische Probleme verhindern die Verbreitung neuer Technologien. Wenige würden privat für die Lizenzgebühren bezahlen wollen, die das Bakom verlangt. Auch würden wenige Geschäftsleute über WLAN-Netze arbeiten, bei denen die Sicherheit nicht gewährleistet werden könnte. Darüber hinaus versuchen die grossen Unternehmen die Verbreitung von spezifischen Technologien zu verhindern, die ihnen in Zukunft eine Gefahr darstellen könnten. Wie man sehen kann ist es ziemlich schwierig für bestimmte Technologien Fuss zu fassen, wenn nicht eine grosse Förderung oder eine grosse Akzeptanz von Seiten der Bevölkerung dahinter steht. Es fehlt eine einheitliche Integration aller einzelnen Individuen um ein Flächendeckendes WLAN-Angebot zu offerieren.

Literaturverzeichnis

[1] Marktentwicklung, ComCom, [online] 2001, www.fedcomcom.ch/comcom/docs/rapport2001/DE/IV_Marktentwicklung.pdf (Accessed 31. January 2005)

[2] Wir über uns, Cablecom, [online] 2004, http://www.cablecom.ch/wirueberuns/ourmarkets-ir.htm (Accessed 31. January 2005)

[3] Bluewin, Bluewin, [online] 2005, http://www.bluewin.ch (Accessed 31. January 2005)

[4] Sunrise, Sunrise, [online] 2005, http://www.sunrise.ch/home.htm (Accessed 31. January 2005)

[5] Vodafone Live - UMTS, Swisscom Mobile, [online] 2005, http://www.swisscom-mobile.ch/prv_asp/prv_home.asp?nid=2230&UserLanguage=D (Accessed 31. January 2005)

[6] Bandbreite, Cablecom, [online] 2003, http://www.cablecom.ch/030718_mmbandbreite.pdf (Accessed 31. January 2005)

[7] Die Historie der Computertechnik, Mac.com, [online] 2005, http://homepage.mac.com/ulrich.behning/Computergeschichte_Preise.html (Accessed 31. January 2005)

[8] Bluewin Internetzugang, Bluewin, [online] 2005, http://de.bluewin.ch/internetzugang/index.php/index_adsl (Accessed 31. January 2005)

[9] Monzoon, Monzoon, [online] 2005 http://www.monzoon.org (Accessed 31. January 2005)

[10] Produkte > WLAN > Hotels, mpol solutions, [online] 2005, http://www.mpol.ch/produkte/network/wlan/wlan_hotel.html (Accessed 18. February 2005)

[11] Faktenblatt des BAKOM über WLAN, BAKOM, [online] 2003, http://www.bakom.ch/imperia/md/content/deutsch/telecomdienste/factsheets/factsheet_wlan.pdf (Accessed 31. January 2005)

[12] FAQ des BAKOM bezüglich WLAN, BAKOM [online] 2003, http://www.bakom.ch/de/telekommunikation/forschung/WLAN/faq/index.html (Accessed 31. January 2005)

[13] Prof. Rolf oppliger, Vorlesung Einführung in die Sicherheitstechnick, 03/04

[14] WEP Unsicherheit,[online] 2004, http://www.isaac.cs.berkeley.edu/isaac/mobicom.pdf (Accessed 31. January 2005)

[15] Verwaltungsgebühren im Fernmeldebereich, [online] 2005, http://www.admin.ch/ch/d/sr/784_106_12/index.html (Accessed 31. January 2005)

[16] Theoretische Schwachstellen von WEP Verschlüsselung, [online] 2005, http://www.cs.umd.edu/w̃aa/wireless.html (Accessed 31. January 2005)

[17] Internet und Sicherheit, Trojaner und Sicherheit, [online] 2005, http://www.trojaner-und-sicherheit.de/tcp_ip_und_internet/vpn.htm (Accessed 31. January 2005)

[18] Rent a VPN, [online] 2005, http://www.rentavpn.de/index.asp?showdoc=14 (Accessed 31. January 2005)

[19] Passwort Check des Kantons Zürich, Datenschutz.ch [online] 2004, https://passwortcheck.datenschutz.ch/doc/goodchoice.de.php (Accessed 31. January 2005)

[20] Cern Security Handbook, CERN, [online] 2005, http://consult.cern.ch/writeups/security/security_3.html#SEC7 (Accessed 31. January 2005)

[21] Bluetooth-Angriffe aus großer Entfernung, Heise online, [online] 2004, http://www.heise.de/security/news/meldung/49907 (Accessed 31. January 2005)

[22] Java-Sicherheit auf Handys ausgehebelt, Heise Online, [online] 2004, http://www.heise.de/security/result.xhtml?url=/security/news/meldung/52321&words=Sicherheit%20Mobiltelefonen (Accessed 31. January 2005)

[23] Mobiles unter Beschuss, Heise Online, [online] 2004, http://www.heise.de/mobil/artikel/50820/0 (Accessed 31. January 2005)

[24] BlueBug, Trifinite.org, [online] 2004, http://trifinite.org/trifinite_stuff_bluebug.html (Accessed 31. January 2005)

[25] BlueSnarf, Trifinite.org, [online] 2004, http://trifinite.org/Downloads/BlueSnarf_CeBIT2004.pdf (Accessed 31. January 2005)

[26] Security für Smartphones, Computerworld Online, [online] 2004, http://www.computerworld.ch/news/0411081120.html (Accessed 31. January 2005)

[27] Crack für WLAN-Verschlüsselung WEP aufgetaucht, Heise Online, [online] 2001, http://www.heise.de/newsticker/meldung/20342 (Accessed 31. January 2005)

[28] AirSnort, AirSnort [online] 2005, http://airsnort.shmoo.com/ (Accessed 31. January 2005)

[29] Symantec Client Security for Nokia Communicator, Symantec, [online] 2005, http://enterprisesecurity.symantec.com/products/products.cfm?ProductID=663&EID=0 (Accessed 31. January 2005)

[30] Symantec Firewall and Antivirus, Nokia, [online] 2005, http://www.nokia.com/nokia/0,,66410,00.html (Accessed 31. January 2005)

[31] Digitales Zertifikat, Wikipedia, [online] 2005, http://de.wikipedia.org/wiki/Zertifikat_%28Informatik%29 (Accessed 31. January 2005)

[32] Vorlesung Informatik und Recht, Universität Zürich, WS 2004/2005

[33] Clayton M. Christensen: The Innovators Dilemma, HarperBusiness

[34] John C. Dvorak, PC Magazine, [online] 2004, http://www.pcmag.com/article2/0,1759,1722125,00.asp (Accessed 31. January 2005)

[35] Disruptive Technology, Wikipedia, [online] 2004, http://en.wikipedia.org/wiki/Disruptive_technology (Accessed 31. January 2005)

[36] Michael E. Porter: Competitive Strategy, ISBN: 0743260880

[37] Wimax Schweiz, Wimax Homepage, [online] 2004, http://www.wimax.ch/wimax.html (Accessed 31. January 2005)

[38] Sonntagszeitung.ch, [online] 2004, http://www.sonntagszeitung.ch/dyn/news/print/multimedia/372667.htm (Accessed 31. January 2005)

[39] Michael Cai, Wi-Fi Planet, [online] 2005, http://www.wi-fiplanet.com/wimax/print.php/3364141 (Accessed 31. January 2005)

[40] WiMax, Teltarif.ch, [online] 2004, http://www.teltarif.ch/i/wimax.html?page=2 (Accessed 31. January 2005)

Chapter 7

New Opportunities for B2B and B2C Services Offered by New Mobile Communication Technologies

Christoph Müller, Dane Marjanovic, Thomas Schönenberger

First, we give an introduction to the most common and most widely used technologies concerning telecommunication. Each technology is discussed focusing on mostly technical issues, showing some insightful information about each technology. A comparisons among the technologies is also presented. After the technical background, the paper investigates also economical aspects, focusing on the mobile commerce concerning topics like value chain modelling and a discussion of possible new business models enabled by mobile communication technologies. The goal is to address economical implications of new mobile communication technologies to the market and the value chain of a company. Afterwards, the paper addresses the future, providing a descriptive but verry insightful view of the future of mobile telecommunication. It shows the future of mobile communication in form of scenarios depicting the world in the year 2010. It shows both economical and social circumstances at that time. These scenarios are derived from technological, social and economical trends we can observe nowadays. Finally, this paper gives a very detailed picture of the mobile communication world showing both present and future circumstances.

Contents

7.1 Introduction

By looking at the present technology standings in mobile communication we witness the beginning of a slow technological shift from the currently deployed 3rd Generation (3G) (including the 2.5th Generation) to a 4th Generation (4G) technological infrastructure. This shft however is to be observed with patience since the changes and improvements will probably not take effect until 2008 or 2010. These improvements will bring new infrastructures, new ways of communicating, new business models, new devices and new opportunities for business to business and business to consumer services.

Regarding the above mentioned consideration the paper focuses on 3 main topics: Mobile Technology, Mobile Market and Forecast into 4th Generation. The Mobile Technology section refers to all present and future technologies and infrastructures. The section starts with a short history review of mobile communication. It then focuses on each technology (standard) in mobile communication, such as HSCSD, GSM, EDGE, etc., in turn.

The second section adresses mobile commerce, its value chain and possible new business models enabled by new technology. Further it makes some considerations about possible new killer applications regarding service providing and usage of new devices. The value chain part is concerns thoughts such as content creation, content aggregation and portal-services.

The third part adresses possible future developments in mobile communication technology. It tryes to make an outlook into 4G technology, 4G infrastructures and the deployment and usage of the menioned. The outlook is concieved as a set of scenarios depicting possible usagepatterns of the 4G technology and infrastructure. These scenarios are based on present trends we can observe in mobile communication and take some considerations from the first two chapters into account.

7.2 Mobile Technology

7.2.1 A short history of mobile communication

The history began in about 1960. For example, in Germany the fist communication net was established in 1958. The so called A1 net was arranged by hand and covered 80% of the area and 95% of the population in Germany in 1970. The A2 and A3 net improved the capacity, but only to a new limitation of 11000 members. In the end there were 600 persons to arrange the connections, that meant 18 members of the services had to pay one worker. So the price was very high. In 1972 the A net was replaced by a new network. But this network was still a network of the first generation. In this network the connections where no longer arranged by hand and a capacity of 13000 has been reached in 1979. The enlargement of 27000 possible members was full in 1986. In the year 1986 a new network was ready to launch. It was a net with analog technology too, but it was much more improved. The transmission was encrypted and the capacity limitation increased to 400000 costumers in Germany. Another innovation was, that the end devices

were portable. The smallest was about 700 grams. Other advantages were the SIM-card, which made it impossible to use the network without a contract and it was not longer necessary for the customer to know the exact location of his/her dialog partner. In Switzerland this chapter of mobile communication was over at the end of 1999. In some areas it may still be available.

In 1982 a working group has been founded by the CEPT (European Conference of Postal and Telecommunications Administrations). The challenge was to design a "'standard for mobile communication"' in Europe. The Name of the group was **G**roupe **S**peciale **M**obile or GSM. Late it was renamed to "'**G**lobal **S**ystem for **M**obil communications"' because the standard was also used in other countries outside of Europe. In 1987 the decision was made to build a digital network and about 1990 the first network was online. In Switzerland the fist digital cellphone was available during the Telecom 91 exposition in Geneva. But two more years were needed for the commercial network. In Germany the first network was online on the 1.7.1992 named D1 and D2.

In 1995 there were two new enhancements in Switzerland. The first was the new GSM specification DCS-1800. This is a specification for new frequencies and expected capacity. But for two years there was no possibility to use both standards together. The second innovation was the Short Message Service or SMS. In the last ten years the number of users got higher and higher. On the other hand the cost of a cellphone decreased, especially with the launch of prepaid-offers. This changed the profile of the cellphone user. At the beginning it was a useful tool for businessmen, but later everybody could use it and it became very popular by young people.

The GSM standard is an open standard. So, there are many different companies producing end devices. But also on the network side are different companies – for example Siemens or Alcatel. GSM is a technology of the second generation. Besides the voice communication there is also data transmission, but the maximum speed is 9600 bit/s. At the moment there are different innovations to get over this limit. The names therefore are GPRS, HSCSD and EDGE, and the improvements of these technologies are channel bundling and a possibility to pay per byte. This is more than the planed second generation was able to. So it is called the 2.5G.

The third generation or 3G comes with the new UMTS standard. In theory the commercial beginning was a long time ago, but there were different problems with the technology and the start shifted again and again. But there are numerous rumours of UMTS cellphones for Christmas 2004. UMTS became very famous in press, but not for its possibilities. It were the billions that have been spent for the frequencies. In Germany, the six winners of the auction paid six billion Euro each, only for the license. So it is not likely, that they will reach the break-even very soon. Another problem is, that there is no application for UMTS. Each demand can be satisfied with the old technology.[2] [3]

7.2.2 Technical overview

The GSM network has a hierarchical structure. The central networks are multiple **S**witching **S**ub**s**ystems (SSS). Some of it take over also the linkage to the external world, link to the

traditional telephone network and to other GSM nets. A SSS contains a distributed data base of the portable radio participants and takes over the authentification of the participants. At a Switching Subsystem there are in each case several **B**asis **S**tation Controls connected. These again steer several **B**asis **T**ransceiver **S**tations, which undertake the task of actual radio transmission to the mobile devices.

The basis stations cooperate using firmly assigned frequencies, which differ from those of the neighbour stations. The entire net is divided in cells, in which certain frequencies apply, in each case, and which are surrounded by other cells, which have likewise own frequencies. A radio cell can have an expansion between 500 m and 35 km depending upon topography and land development. While planning a net, extensive measurements have to be accomplished, so that with as few basis stations as possible, the best possible net cover can be reached.[2] [3]

7.2.3 GSM

GSM works with paired frequency ranges. For the connection of the mobile equipment with the basis station the range is 890 - 915 MHz, for the opposite direction the range is 935 - 960 MHz. Both ranges have thereby a distance of 45 MHz. With DCS- 1800 the ranges are 1710 - 1785 MHz and 1805 - 1880 MHz. The frequencies are divided into channels, which have in each case a range of 200 kHz. The 900 MHz volume has thereby 124 channels.

Each channel is splited using the time division multiple access method (TDMA) in 8 time slots. In each time slot there are 142 bits transferred, what is called a "'burst"'. Eight of these bursts are a frame in TDMA. The basic station transfers such frames in continuous sequences. A mobile equipment that has its information in the sixth slot, has to receive data only in his slot for recieving a continuous data stream. The time division multiplexing is used also for the transmission from the mobile station to the basis station. Each mobile equipment keeps assigned for it a time slot and a transmitter frequency firmly, i.e. each frequency has up to eight mobile devices.[2] [3]

7.2.4 HSCSD, GPRS and Edge

HSCSD

HSCSD stands for High Speed Circuit Switched Data and makes higher data transmission rates possible for data, as several GSM channels are hooked up. In addition only a change of software has to be made. On the hardware side the necessary devices are already present in the mediating net. Differently the thing looks with the terminals. Since for mobile phone no software updates are intended, new devices must again be bought. HSCSD increases the data rate per channel from 9600 Bit/s to 14400 Bit/s, as error correction is less complex. However the risk of transfer errors rises. If the connection is not good enough, it must be shifted back again to 9600 Bit/s.

Theoretically all eight bursts of a TDMA frames could be occupied with HSCSD, what would result in a data rate of 115,2 kBit/s. Thus a participant could use an entire carrier frequency for itself, what could lead to the impairment of the availability for other participants. Therefore the network carriers limit the maximum number of provable channels at the same time at 3 or 4 (43.2 kBit/s or 57.6 kBit/s).

The network carriers configure their nets normally in such a way that HSCSD channels are made available only if these are not needed for voice connections. Therefore the user can also be cut off from some channels although he has already been assigned to them during a session. Besides also an asymmetric connection is possible. The user receives then e.g. two channels to the receipt, but only one for sending data. The transmission costs are measured with HSCSD by time and by the occupied channels - with four channels one pays therefore also the quadruple price for the connection.

GPRS

With **G**eneral **P**acket **R**adio **S**ervices, a permanent Internet entrance for mobile devices has been introduced. A contract with the network carrier is concluded – which becomes thereby at the same time the Internetprovider. The data is being transmitted with GPRS packetwise, whereby the network carrier decides in each case, how many data channels are available for the transmission. Under ideal conditions, i.e. if the carrier frequency is otherwise unused, GPRS can occupy all eight channels. The basic data transmission rate can be varied in four stages, which are dependent on quality of transmission: coding Scheme 1 supplies the highest transmission reliability with only 9.05 kBit/s, coding scheme has 13,4 kBit/s, coding scheme 3 15.6 kBit/s and coding scheme 4 has 21.4 kBit/s.

Edge

Edge (Enhanced Data of advice for GSM evolution) will be based on an advancement of the GSM standard and would require at least a part of the exchange technique with the portable radio network carriers. Considering the investments in UMTS it is questionable whether still another operator will make this expensive intermediate step.

Edge is as extension of HSCSD and GPRS. It would be named ESCD or EGPRS. Both kept thereby their fundamental characteristics, make however data rates available from theoreticaly upto 473,6 kBit/s, whereby in practice data rates are approx.. 170 kBit/s. Edge would permit beyond that also a higher extent of utilization of the existing frequencies for normal voice connections. Like that more calls would be obtainable, without making the radio cells smaller and setting up additional basis stations.

Which for what?

HSCSD is very well suitable, in order to develop direct data links, thus concerning speed. GPRS serves besides it rather as an Internet entrance with always- on- functionality.

That can be on the one hand for the binding of field representatives to the LAN of the center useful, requiring however a secure connection to prevent monitoring, e.g. by the employment of VPN Routers at both terminals. Both variants are momentarily still horring high prices. In order to surf with approximately ISDN speed in the Internet, one pays with HSCSD momentarily still four times the price that one would have to pay for a mobile phone discussion at the same time. With GPRS the fee amounts to approx.. 1 cent/kByte - an average web page with 50 KB strikes thus already with 50 cents.

The network carriers momentarily still look for services, which make GPRS and HSCSD interesting, also in connection with the to a large extent defamed WAP. In addition they want to use e.g. the locatableness of mobile phones in the GSM net. This can bring so-called location to Based services dependent on the location of for example the local weather forecast, the program of the nearest cinema or the good deals of the surrounding business on the mobile phone display. This can take place also via push services, i.e. one does not have to fetch the information, but becomes it (if one with it agrees) directly on the mobile phone to transfer.[2] [3]

7.2.5 UMTS (Universal Mobile Telecommunication Service)

Since 1999 it is working on the technical specifications for the portable radio net of the third generation. In addition the ETSI (European Telecommunications standard of institutes), ARIB (Japan) and T1P1 (North America) have joined the project group "'3GPP"'. To a completely uniform standard they did not bring it, nevertheless - the respective interest was too large to make UMTS the technical extension of the portable radio standard of the second generation usual in the respective homeland region. Thus also UMTS will make world-wide roaming possible despite large standardization presumably only with special, complex and thus expensive terminals. Also with the designation no complete agreement prevails: In North America the mobile net of the third generation becomes designated as "'IMT-2000"'.

Licenses and frequencies

With the contracting methods for the UMTS licenses also the frequency ranges planned for it were distributed. Internationally by the International Telecommunication Union for UMTS the frequencies were planned 1900 - 2025 MHz and 2110 - 2200 MHz. Excluded of it however are the frequencies 1980 - 2010 MHz and 2170 - 2200 MHz for applications of satellites.

Network structure

The UMTS net resembles very strong the GSM net. The UMTS core net corresponds to a large extent to the group of the switching centers (SSS) in the GSM. The characteristic is however that it concerns thereby a package-obtaining core net on the basis of IP.

Transmission technique

The medium access in the UMTS is based on a frequency band spreading named "'WCDMA"' (Wideband Code Division Multiple Access). The transferred data with a higher-frequency, to pseudocoincidental bit sequence overlay. Thus neighbouring cells can use the same frequencies – by the band spread a terminal can repair the original data from the receive data by demodulation using the correct bit sequence. The maximum data transmission rate on a channel amounts up to 2 Mbit/s, however a participant has it for itself alone only so long, as he doesn´t move faster than 6 km/h and is not further away from the next basis station than 500m. Otherwise the data transmission rate sinks on maximum 384 kBit/s, 144 kBit/s by the network carrier is guaranteed.

Similarly as with GPRS one divides the existing transmission capacity with the remaining participants, who are in the same cell. The allocation will take place after service classes, i.e. by a higher tariff one can buy oneself priorities in the net. Exactly the same as with GPRS the devices are always on-line, if they announced themselves in the net.

Since all cells with the same frequencies work, the net capacity can be increased simply by adding further base stations. The surrounding cells then automatically make their ranges smaller. This is used particularly with the structure of the net: First a few base stations in large distances are set up, which cover a large range in each case. In population centres gradual further basis stations are then developed. Thus rising user numbers are always ordered the same service. The start-up of the first UMTS nets is intended for 2004. These will first address mainly business customers. The break-through in the Consumerbusiness could be the year 2010.[2] [3]

7.2.6 SMS, MMS and WAP

SMS

The message payload is 140 bytes: either 160 7-bit characters, 140 8-bit characters, or 70 2-byte characters in languages such as Chinese, Korean, or Japanese. This does not include routing data and other metadata, which is additional to the payload size. Short message services are developing very fast throughout the world. By mid-2004 texts were being sent at a rate of 500 billion messages per year. At an average cost of 10 cent per message, this generates revenues in excess of $50billion for mobile telephony operators and represents close to 100 text messages for every person on the world. Growth has been rapid; in 2001, 250 billion SMS were sent, in 2000 just 17 billion. SMS is particularly popular in Europe, Asia and Australia. Popularity has grown to a sufficient extent that the term texting (used as a verb meaning the act of cell phone users sending SMS text messages back and forth) has entered the common lexicon.

The most frequenter texters are found in south-east Asia. In Singapore, hundreds of messages can be sent per month for free, after which messages cost between SGD 0.05 and SGD 0.07 each to send. The same pricing format is followed in the Philippines where the average user sent 2,300 messages in 2003, making it the world's most avid texting

nation. Europe follows next behind Asia in terms of the popularity of texting. Users in Spain sent a little more than fifty messages per month in 2003. In Italy, Germany and the United Kingdom the figure was around 35-40 texts per month. In the US, however, the appeal of SMS is even more limited. Although a SMS usually costs only USD 0.05, only 13 messages were sent by the average user in 2003. The reasons for this are various – many users have unlimited "'mobile-to-mobile"' minutes, high monthly minute allotments, or unlimited service. Moreover, "'walkie talkie"' services offer the instant connectivity of SMS service and are typically unlimited. New SMS services offer automated "'alerts"' sent on a regular base giving news, weather, financial information, sporting event scores, and other information.

SMS is also increasingly being used for "'real-world"' services. For example, some vending machines now allow payment by sending an SMS; usually, the cost of the item bought is added to the user's phone bill. A few mobile phones allow long SMS messages to be sent. This is accomplished by breaking up the long message into shorter messages and adding some code indicating that the messages should be strung together on the recipient's phone. It should be noted, however, that this does not count as just one SMS.

Several telecommunication carriers have recently started offering so called premium rate short messages, which through higher pricing and revenue sharing allow companies to be paid for their services by sending a short message. This is also becoming increasingly popular, but problems arise when the premium pricing is not advertised.

An increasing trend towards spamming cell phone users through SMS has prompted cellular service carriers to take steps against the practice, before it becomes a widespread problem. No major spamming incidents involving SMS have been reported as of October 2003, but the existence of cell-phone spam has already been noted by industry watchdogs, including Consumer Reports magazine.[1]

MMS

Multimedia Messaging System is the logical evolution of the Short-Message System. MMS-capable mobile phones enable subscribers to compose and send messages with one or more multimedia (digital photos, audio, video) parts. Mobile phones with built-in or attached cameras, or with built-in MP3 players are very likely to also have an MMS messaging client.

There are some interesting challenges with MMS that do not exist with SMS. The most interessting one is the content adaptation. Multimedia content created by one brand of MMS phone may not be entirely compatible with the capabilities of the recipients' MMS phone. In the MMS architecture, the gateway is responsible for providing content adaptation, if this feature is enabled by the mobile network operator. When content adaptation is supported by a network operator, its MMS subscribers enjoy compatibility with a larger network of MMS users than would be available otherwise .

MMS was originally developed within the Third-Generation Partnership Program (3GPP), a standards organization focused on standards for the UMTS/GSM networks. Today MMS has been deployed world-wide and across both GSM/GPRS and CDMA networks.[1]

WAP

The primary language of the WAP specification is WML, the Wireless Markup Language, which has been designed from scratch for handheld devices with phone-specific features following the XML guidelines. The new version of WAP, WAP 2.0, is a re-engineering of WAP using XML. Some observers predict that this next-generation WAP will converge with, and be replaced by, true Web access to pocket devices. XHTML Basic, a subset of XHTML, is made to work in mobile devices.

WAP Push
WAP Push, available since WAP 1.2, has been incorporated into the specification to allow WAP content to be pushed to the mobile handset with minimum user intervention. A WAP Push is basically a specially encoded message which includes a link to a WAP address. A WAP Push can be delivered over WAP bearer or SMS bearer. When receiving a WAP Push, a WAP 1.2 or later enabled handset will automatically give the user the option to access the WAP content.

In this way, the WAP Push directs the end user to a WAP address where particular content may be stored ready for viewing or downloading to the handset. The address could be a simple page or multimedia content (e.g. polyphonic ring tone) or a Java application. Using WAP Push, one can make it easier for end users to discover and access new mobile services.

Failure?
WAP was intended as a mobile replacement for the World Wide Web. However, its idiosyncratic protocols cut users off from the true HTML / HTTP Web, leaving only native WAP content and Web-to-WAP proxy content available to WAP users. WAP's charging model, where users have to pay by the minute regardless of the amount of data received, has also been criticized.

WAP was hyped at the time of its introduction, leading users to expect WAP to have the performance of the Web. One telco's advertising showed a cartoon WAP user "'surfing"' through a Neuromancer-like "'information space"'. In terms of speed, ease of use, appearance and interoperability, the reality fell far short of expectations. This led to the unkind, but widely used phrases: "'WAP is crap"', "'Worthless Application Protocol"', "'Wait And Pay"'.

The main reasons for the failure of WAP were price and closedness. Even though GPRS made WAP cheap, and cell phone operators opened their gateways to access all of the Internet, WAP did not quite take off.

Success?
However, WAP has seen huge success in Japan. While the largest operator NTT DoCoMo has famously disdained WAP in favor of its in-house system i-mode, rival operators KDDI (au) and Vodafone Japan have both been successful with the WAP technology. In particular, J-Phone's Sha-Mail picture mail and Java (JSCL) services, as well as au's chakuuta/chakumovie (ringtone song/ringtone movie) services are based on WAP. After being shadowed by the initial success of i-mode, the two smaller Japanese operators have been gaining market share from DoCoMo since spring 2001.

Korea is also leading the world in providing advanced WAP services. WAP on top of the cdma2000 network has been proven to be the state of the art wireless data infrastructure. According to the Mobile Data Association, June 2004 has seen a considerable increase of 42% in its recorded number of WAP pages viewed compared with the same period in 2003. This takes the total for the second quarter of 2004 to 4 billion.[1]

7.2.7 iMode

NTT DoCoMo's i-mode is a wireless internet service for i-mode mobile phones using HTTP, popular in Japan and increasingly also elsewhere. It was inspired by the drawbacks of WAP being discussed at the time, and a rough concept aimed for businessmen introduced to DoCoMo by McKinsey in 1997. After a 2 year period of developments at NTT DoCoMo, i-mode was launched as a mass-market product in Japan on February 22, 1999. i-mode is a mobile Internet service which as opposed to the mobile specific WAP standard which utilizes WML, instead uses standard Internet HTML, C-HTML, as well as DoCoMo proprietary ALP (HTTP) and TLP (TCP, UDP) protocols. It became a runaway success because of the use of de-facto Internet standards, its well designed services and business model. It uses C-HTML (for "'Compact HTML"') markup language, a subset of HTML. Later on NTT DoCoMo's collaboration with Sun Microsystem resulted in DoJa i-appli, an extension of CLDC (not compatible with J2ME/MIDP). There are some 3000 content-providing companies for i-mode.

As of August 2004, i-mode has 43 million customers in Japan, and over 3 million in the rest of the world. i-mode is being provided world-wide through DoCoMo's partners through a licensing scheme through mobile operators in countries like Germany, the Netherlands, Belgium, France, Spain, Italy and so on.

i-mode is now a global ecosystem where the operators design the compatible handsets, give strong editorial and usability rules to content providers, and propose an open business model to them. i-mode uses open standards as technologies: a light version of HTML is used for producing content, the i-mode mail is interoperable with Internet e-mail and images and sound formats are the ones used on the Web. i-mode users have access to various services, e.g. e-mail, sports results, weather forecast, games, financial services and ticket booking.

i-mode phones

Some typical features include so-called shell model, large displays (250x 320 pixels) and in many models, a display on either side. Additionally the phones have many extra features, e.g. a megapixel digital camera.[1]

7.2.8 Microbrowser

A microbrowser is an internet or web browser designed for use on a handheld device such as a PDA or mobile phone. Microbrowsers have small file sizes to accommodate the low

memory capacity and low-bandwidth of wireless handheld devices. Essentially it is a stripped-down web browser. Sometimes it is referred to as micro-browser, mini-browser or minibrowser.

Underlying technology

The microbrowser usually sets up the cellular networks themselves and translates HTML from the internet into WML based on HDML which is then stored on the server. WML and HDML are stripped-down formats suitable for transmission across limited bandwidth, and wireless data connection called WAP.

Pioneers

The so-called microbrowser technologies such as WAP, NTTDocomo's i-mode platform and Openwave's HDML platform have fuelled the first wave of interest in wireless data services.

Small Screen Rendering Limitations

As mentioned, not only do microbrowsers need to be small in file size, the display screen is also much smaller. Extreme care and meticulous detail must be considered in displaying HTML information onto such a small screen. Bandwidth is also extremely limited and so is the stability. Connections get cut off as with ordinary cell phones and PDAs that are wirelessly connected.[1]

7.2.9 PDA OS

PalmOS

Palm OS is an operating system made by PalmSource, Inc. for personal digital assistants (PDAs) manufactured by various licensees. Contents

History
Palm OS was originally developed on the original Pilot PDA by US Robotics. Version 1.0 was present on the original Pilot 1000 and 5000 and version 2.0 was introduced with the PalmPilot Personal and Professional. Version 5.0 was introduced with the Tungsten T and was the first version released to support ARM devices. Described as a stepping stone to full ARM support, Palm apps are run in an emulated environment called the Palm Application Compatibility Environment (PACE), decreasing speed but allowing great compatibility with old programs. New software can take advantage of the ARM processors with ARMlets, small units of ARM code. It was also roughly this time when Palm began to separate its hardware and OS efforts, eventually becoming two companies,

PalmSource, Inc. (OS) and palmOne, Inc. (hardware). Further releases of PalmOS 5 have seen a standardised API for hi-res and dynamic input areas, along with a number of more minor improvements. Palm OS 5.2 and 4.2 (and later) also feature Graffiti 2, due to the loss of a patent infringement lawsuit with Xerox. This is based on Jot by CIC.

PalmSource, Inc. released Palm OS Cobalt (also known as Palm OS 6) to licencees on December 29th, 2003. This is to be the completion of the migration to ARM devices, and allow ARM native applications along with improved multimedia support. Currently no hardware products yet run Palm OS Cobalt. Built-in applications for Palm OS are Address, Calculator, Date Book, Memo Pad, Note Pad and To Do list

Add-on applications
There are many successful Palm add-on applications. As of August 2003, there are more that 19,000 add-on applications available for the Palm platform, which have various licensing types, including open-source, freeware, shareware, and traditional commercial applications.

Screen sizes
Multiple resolutions are also supported. The original Palm used 160x160 pixels. Some early third party handhelds could collapse the graffiti area for 160x250 pixels. Hi-res is now available with 320x320 pixels and some handhelds like TapWave Zodiac, Sony's Clie range or the PalmOne Tungsten T3 and Tungsten T5 come with 320x480 Hires+ displays with virtual graffiti areas. PocketPCs typically have screen resolutions of 320x240 pixels, though newer models support 640x480 pixels.

Windows CE

Windows CE (sometimes abbreviated WinCE) is a variation of Microsoft's Windows operating system for minimalistic computers and embedded systems. Windows CE is a distinctly different kernel, rather than a 'trimmed down' version of desktop Windows. It is supported on Intel x86 and lookalikes, MIPS, ARM family, and Hitachi SuperH processors. Windows CE is optimized for devices that have minimal storage - a Windows CE kernel may run in under a megabyte of memory. Devices are often configured without disk storage, and may be configured as a 'closed' system that does not allow for end user extension. Windows CE conforms to the definition of a real-time operating system, with a deterministic interrupt latency.

Microsoft has stated that the "CE" is not an intentional acronym, but many people believe CE stands for "'Consumer Electronics"' or "'Compact Edition"'. The first version, known during development under the codename "'Pegasus"', featured a Windows-like GUI and a number of Microsoft's popular applications, all trimmed down for smaller storage, memory, and speed of the palmtops of the day.

Since then, Windows CE has evolved into – according to Microsoft's documentation – a component-based, embedded, real-time operating system. It is no longer only targeted at hand-held computers. Many platforms have been based on the core Windows CE operating system, including Microsoft's Handheld PC, Pocket PC, Pocket PC 2002, and Smartphone 2002. Versions of Windows CE were even made to run on the Sega Dreamcast.

EPOC

EPOC is a range of Operating Systems developed by Psion for portable devices, primarily PDAs. Releases 1 and 2 (unofficially known as EPOC16) refer to operating systems used in Psion's "'SIBO"' devices, which have in common an 8086-family processor and a 16-bit architecture. The first EPOC16 device was shipped in 1989, and as of November 2004, the Workabout mx is still in production. Releases 3, 4 and 5 of EPOC (unofficially known as EPOC32) are a brand new 32-bit operating system. Release 3 first appeared with the release of the Psion Series 5 in 1997.

During development of EPOC32, Psion planned to license EPOC32 to third-party device manufacturers, and span off its software division as Psion Software. One of the first licensees was the short-lived Geofox, which halted production with less than 1,000 units sold. Ericsson marketed a rebranded Psion Series 5mx called the MC218, and later created an EPOC-based smartphone. Oregon Scientific also released a budget EPOC device, the Osaris.

In June 1998, Psion Software became Symbian, a major joint venture between Psion and phone manufacturers Ericsson , Motorola and Nokia. As of Release 6, EPOC became known simply as Symbian OS.[1]

7.3 Mobile Market

7.3.1 M-Commerce vs E-Commerce

This chapter gives a definition of the term mobile commerce in relation to electronic commerce, electronic business and mobile business. The goal of this part is making you understand what the fine differences between these terms mean and what definition would be appropriate for m-commerce. Entered into the Google search-form, the string "'m-commerce"' gets the following definitions:

Unisys.com – Mobile commerce. The use of mobile devices, such as wireless phones to conduct e-business [4]. E-business they define on the same site by: the use of both technology and new business strategies to conduct business online. E-Business provides a sales, marketing and information channel in the internet economy [4]. E-Commerce they call the use of technology to conduct financial transactions online – such as the buying and selling of products and services – usually via the internet. The financial-transactions side of E-Business [4].

indbazaar.com – A term referring to Mobile Commerce, a hybrid of e-commerce. Mobile commerce is effectively the ability to conduct monetary transactions via a mobile device, such as a WAP enabled cell phone. M-commerce is seen as the Holy Grail of the wireless device market [5].

mcgraw-hill.com – (M-commerce) allows you to use wireless devices such as cell phones or PDAs to buy and sell products and services [6]. That's what the public means about

this new, fast growing part of the internet economy. But what do the mobile device manufacturers mean? A good and generic definition of m-commerce as defined by Nokia is: Electronic commerce in which transactions are made using a wireless device and data connection. Mobile commerce can be used to buy things, make banking transactions, and receive information from web sites related to sports, stock, weather, and news, for example [7].

Siemens Program and System Engineering understands by the term mobile commerce the following: Mobile commerce (or m-commerce) is the summary term used for all kinds of transactions – in particular in the context of buying and selling - conducted between business partners and based on mobile electronic devices (mobile phones, PDAs, ...) and mobile radio networks. The fields of application for m-commerce range from mobile online shopping to mobile banking and brokering, reservations and bookings to mobile video games, and a lot more [9]. This definition, despite being very detailed, nearly hits the bull's-eye. Another definition has been penned by the T-Mobile-Stiftungsprofessur für M-Commerce (www.m-lehrstuhl.de) at the University of Frankfurt: The utilization of mobile devices and communicating and interacting possibilities for mobile applications and business areas[1][8].

Having some definitions, we can now put m-commerce in its proper place in the field of e-business. Is it just a part of e-business? Or of e-commerce? In the broader sense m-commerce means the something similar to e-business, but only on mostly wireless, mobile devices. E-Business is meant in the sense of any information system or application that empowers business processes [1]. In the narrower sense it means mobile e-commerce as buying, selling, marketing and servicing of products or services made through mobile devices [1].

7.3.2 Market Overview

The mobile market is one of the fastest growing markets in the last years. This chapter will present some facts and figures about the mobile market in different regions and about the internet. In 2004 the ratio between fixed net telephone subscribers and mobile phone users has been inverted, what means that there are now more subscribers for cellular phones than the 1.18 billion wired-phone users. Latest statistics say that there are 1.48 billion mobile phone subscribers all over the world. This is double the users in the year 2000.

What is the reason for this fast growth in the last four years? To answer this question we have to segment the mobile market into the three major regions: North America, Europe and Asia Pacific.

North America: Though leading in internet economies the United States and Canada have not got leading in wireless web services such as WAP. The only widespread networking tools are so called RIM-wireless handhelds that provide an always-on service for sending and receiving corporate e-mail.

[1]Translated from the german definition: "'Die Nutzung mobiler Geräte und Kommunikations- und Interaktionsmöglichkeiten für mobile Anwendungen und Geschäftsfelder"'

Europe: With its 2nd generation network – GSM – that accounts for over 64% of the world's wireless market [10], Europe was the long time leader in mobile services. With Short Message Service as its killer application and about 500 billion Messages sent in 2004, Europe is still in a strong position in the development of mobile services. But Services over the wireless internet haven't achieved acceptance till now.

Asia Pacific: This region is the reason for the fast growth of mobile services and with it the growth of mobile user penetration. NTT DoCoMo with its i-mode service in Japan can tell us one of the biggest success stories for the wireless industry. Started in 1999 Japan's largest mobile operator has become the worlds leader in wireless services like mobile banking, reservation systems etc. with about 42 million i-mode subscribers (November 2004). Also to this success story belongs the start-up of FOMA, Japan's 3rd generation technology, in October 2001 which introduced high volume and high speed data transmission that enabled new services such as video telephony [11]. Another reason for the growth is the penetration in China that counts more than 310 millions of mobile users, what is more than the population of the USA (by July 2004: 293 millions).

This fast growth is paired with a huge increase in revenues in the mobile sector. By 2003 international revenues reached 488 billions Swiss francs [12]. This facts show that by developing of mobile commerce solutions a huge public can be addressed with helpful services in daily life to get more and more revenues. Juniper Research published a new study that predicts the global mobile commerce (excl. m-entertainment) to become a 40bn industry by 2009, fuelled by a growth in micropayment volumes. This will particularly be pushed by the purchase of tickets using the mobile phone [16].

Other than the internet part of e-business the new mobile commerce will address to a broader public, because there are more people that already own a cellular phone, than an internet-connected computer: All over the world, 699 millions people are internet users and only 102 have broadband internet connections compared to 1.48 billions mobile subscribers. Key players in mobile commerce are NTT DoCoMo with their iMode-service having over 42 millions of mobile data users and 7 millions of 3G subscribers and KDDI with their EZ-Web-portal having only 17 millions of mobile data users but almost all of them using 3G devices in Japan.

7.3.3 Content Creation

The next three chapters will describe the value chain of content production and presentation for mobile markets. First there is the source of all content, the firm that creates messages, video- and audio-clips and graphics. This step is here called content creation and will be described in this chapter. Second there must be someone to aggregate all those medias and order by topic, region etc. We call this part content packaging, handled in the next chapter. The third step is to present all that content to the public. These content presenters are mostly mobile portals. More on that part you will read in chapter 7.3.5. The focus in content creation is on creating digital material such as audio, video and textual information. We can call it, getting the raw material.

The differences between e-commerce- and m-commerce-content are not major. Although there are some restrictions in format and size, the type of content typically for wireless Internet is the following:

Text: Most information is and will in future also be in textual format. A reason for this is, that most information contains facts and figures that can cheaply be relayed in text, another, that text is the best media for displaying on the small screens of mobile devices. Text information could be news, stock prices, product descriptions or restaurant addresses.

Audio: The voice-service is still the main usage of the mobile phone and will also be in future, when prices for telephony will further decrease. Other services like wireless internet radio or downloading music files will become popular.

Graphics: Not only for modifying and customizing the users mobile phone with wallpapers and screensavers, graphics are helpful for services like e.g. weather channels. This content can easily be created because most mobile devices can interpret the jpeg-file format.

Video: Today we know animated graphics from MMS-weather- and similar services. But with the broadband wireless internet, which comes in 3rd generation, there will be wireless TV, video-files-download and other video-services.

Such content can easily be modified, consumed repetitively by the same or other users, and is fast and cheap to reproduce [13]. These formats evocate new issues such as time-dependency – information for a brokerage-service is time-dependent, while an online-dictionary is more or less independent. Another issue is pricing: how each service is being charged. The easiest way to charge for value-added mobile services would be by taking a share in the call revenues, but this solution is out-dated and since there is a strong competition between the mobile operators, prices for calls and calling revenues are decreasing.

There are also emerging problems by creating content because digital content is still limited on the mobile internet. Mobile Devices support formats that differ from standard internet formats. Leading companies in content creation CNN, Reuters, Consors and Webraska (traffic news) are providing data to portals and network providers, as well as developing their own portals (vertically integrated, see also Market Making, Chapter 7.3.5).

7.3.4 Content Packaging

The part of content packaging is a very central one in the value chain of mobile content. Millions of messages are being created today, but not every portal wants to bring every message to the users display. For that reason there are content packagers, which will be further described in this chapter. The value added in this part of the value chain is in delivering content in the most appropriate package. The process covers aggregating and transforming information for distribution to wireless devices. In wireless portals, the

content included should be filtered, customized and aggregated for the specific topics, locations, time and sometimes for each customer individually. The following example gets close to what content packaging is.

A regional travel-portal needs following information to best inform its customers:

- Location based railway schedule with real-time-information
- Traffic information including traffic jam and roadwork information
- Hotel information and booking system
- Restaurant information
- Map of the region
- Local weather

All these information could be packaged by a content aggregator, which does this for multiple regions (centralization) or for one specific location (decentralization). Other processes related to this showcase, like showing the traffic stations of the schedule on a map, are also a part of the job of content aggregator firms.

7.3.5 Market Making

The last part of mobile content production and presentation is the one that mostly influences the mobile market. Therefore we called it market making. Portals are the key for the door between business and consumer. But not only in the business to consumer area portals are the intermediate: in business to business interfaces more and more portals are used to link firms for short- or long-term contracts too. Portal comes from the Latin porta that means doorway or gate: the door to the internet, where information is being aggregated; the intermediate between the demander and the supplier. Goal of a portal is to become the user's prime supplier for web-based information and success of a portal particularly lays in two factors: ease of use and delivering the right information in the right moment, called the value-for-time proposition [14]. Portals play a powerful role in providing access to the mobile internet because users want to have one site where they can get their information tailored to their interests by one click.

Such portals in mobile internet should include services in the following range:

- Personalised content and alerts
- Personal information management (PIM)
- Location-specific information
- Communication

These Services will now be described and later be picked up in chapter 7.3.6, where customer needs will be investigated and possible killer applications will be listed. Communication services include all data-processes used to communicate between two or more users. Such services are for example e-mail or voice-mail. But also upcoming high potential applications like video conferencing and instant messaging are in the range of communication. Personalised content is meant in the sense of news tickers – till now better known in SMS-services – and weather information and alerts for stock prices, betting and auctions. These are only some examples in a wide area of personalised information and useful alerts. Especially alerts have a great potential in the world mobile market.

Personal information management (PIM) is the relief for the filofax. Included services are: calendar, to-do-list, address book, memo pad and other supporting functions. Location specific information applications will best help mobile commerce to get the breakthrough. Such information could be traffic reports, nearest automated teller machine (ATM), hotels and restaurants bookings in the current environment of a person.

All this information lets the portal-provider customize the user's portal site in deep detail, if he has enough information about each user's habits or it lets the user customize the portal by his own. And that is what has to be done better on the wireless internet than on the wired-one, because the small displays of mobile devices cannot hold as much information as a normal webpage. Also the restricted typing and navigation interfaces deter the user from typing long messages and clicking through many sites. While a webpage has 25 links in average, a WAP-Page has about 5 links in average [13]. This restraint makes it indispensable to filter information before provided to the user with a view to get the better *value-for-time* proposition. One key success factor for portals is to hold a large customer base. Below this given point, key players in portal-services can be derived easily. In the strongest position are mobile operators, because they provide the technical infrastructure and thus the last mile to the user. Main advantages over other portal players are the existing billing relationship, substantial subscriber data and the actual location of the subscriber [14].

Mobile operators can constrain the latitude of the mobile internet presented to the consumer and charge a mark-up on services provided by other firms. Such a scenario has already been reality in Japan: Users of the iMode service had only access through the iMode-portal on sites approved by NTT DoCoMo for uniqueness and usefulness. So the Japanese operator was able to take a 9% handling charge. But in 2001 NTT DoCoMo opened up iMode site selections and included also unofficial sites. Virtual operators – mobile operators, which have access to one or more mobile network operators – are also ready to have their portals for their own customer-base. But their position is not as strong as the one of physical mobile operators, because they have little influence on the extension of the network. They only maintain the contract and billing relationships with customers.

Another strong position is being hold by technology vendors; they equip mobile devices with actual technology and can load initial software or bookmarks onto the devices before being sold to customers. Technology vendors often work together with operators to further improve their strong positions. Traditional web portals like Yahoo!, AOL and Excite also try to become established in the mobile internet, but their position is much weaker than the operator's, because they have less information about the user, in the worst case only

the e-mail-address, while an operator can even determine the customers position. That's why most web portals try to cooperate with a regional mobile operator.

Last but not least there are also new players with their business models mostly basing on proprietary technology, which can facilitate the process of content providers. One chief issue in starting as a new player is the decision of marketing under the brand of the operator or under their own brand. But mobile providers are less interested in outsourcing services like portals because of heavy losses in revenues by doing so.

7.3.6 The Killer-Application

Now that we have analyzed the mobile market and its value chain, we want to know, which application will become the killer-application of 3G. Or if there will be any killer-application at all. To know which application will make it, we first have to have a look at the different customer segments. In the business to consumer (B2C) market there are three main segments. First, there are teenagers of 18 years and younger, which are particularly interested in mobile entertainment, such as ringtones, mp3-music and tv-shows. Second, students of 18 to 25 years, engaged in research topics and seeking for information, are also one main target group of mobile content providers and operators. And third there's the main public for news, stock and weather information and all-time availability through mobile communication, the young business people.

Also in the business to business (B2B) market, three main segments are apparent. The first is sales-driven, such as manufacturing companies and banks and needs mobile workforce in the sense of calendar-, e-mail- and groupware-functions. The second is service-driven, for example consultancies and system houses, looking for mobile customer relationship management (CRM) containing mobile sales- and service-applications. And the third is logistics-driven; taxi companies or courier services, seeking for better supply chain management (SCM) possibilities, such as fleet management and track & execute solutions [14].

To evaluate possible killer applications we now have to have a look at the key improvements in these new mobile services. The main improvement is that more and more location-based services will evolve and content providers will offer location specific information. Personalisation is another improvement of future services: portals will filter the information to the user's interests and provide the information he desires accessible by as few clicks as possible. Another progress is immediacy, which enables us to access almost all information in every moment and anywhere we'd like to. This leads to a service that is all time available even outside normal working hours.

Although we have now evaluated some applications and the key improvements in mobile markets we still cannot ascertain the killer-application of the future. But as we know, SMS is the killer-application of the past years, another communication application could become a frequently used service: perhaps it will be instant messaging or something like the availability-service "'Presence"' which most Nokia mobile phones are already equipped with. Another possible market for publicity could be m-advertising, such as sponsored alerts and mobile promotion. Yet another high potential service is the so called m-tailing;

ticketing services as the mobile wallet of Nokia described below, booking services for hotels and restaurant reservation systems.

7.3.7 Pricing

Until now we have learned which services will have the best chances for success in the mobile market. Now it's time to have a look at how the revenues are being collected. This chapter shows how to build price models for all services with mobile devices. There are different ways to charge for mobile services like phone calls, multimedia messaging, downloads or surfing on the internet. Here will also be described which price model matches best for which service and which media.

In the mobile world we best know pricing by time, because every time we make a phone call, the duration of the call will be charged. Sometimes there are minutes included in the fixed monthly fee, but after those free minutes charging by time continues. When the (wired) internet was only with analogue dial-up-lines, it has also been (and still is being) charged by time. Later some ISPs introduced flat-rate price models or price models with monthly free hours included. Then new, faster broadband cable- and DSL-lines where available and had first been charged by the transmitted data volume. When more and more broadband-connections got cheaper, they where only charged with a fixed monthly fee for unlimited access. This chronology could also come true for the mobile internet. An overlook of the world's mobile services and their price models proofs this.

In European countries connections for WAP-surfing with GSM-technology – the oldest way of data-connections – are being charged by time. The newer GPRS-technology gives the possibility to charge by data volume. But there are also other data services that are charged by a monthly fee or included in the subscription, e.g. unlimited access to an e-mail-account (in most cases only the e-mail-account provided by the mobile operator, like in Switzerland by Orange). In Japan, NTT DoCoMo's iMode-service first introduced charging by service. The difference to other price models is that for the first time a value can be given to different kinds of information. That's like in the internet, where additional information services can be paid by credit cards. How these services can be paid will be explored in the next chapter. In Europe such services are SMS- or MMS-information-services like weather information, news and sports-information, ringtones, games and horoscope, each of them being charged with a different price.

One widespread pricing model is charging by location. In the last years more and more mobile operators offer a call service, whereby calls from different locations have different rates. In Switzerland, a similar service by Sunrise is called Sunrise myzone where one place (one antenna) can be specified to be the home-place (or the working-place) and from where calls are cheaper, than from any other place. This model is predominantly designed for people, who do not subscribe for a fixed phone connection at home. But from an operator's view this model is also a powerful tool to better divide customer segments if applied to other location-based services.

Not every price model is reasonable for every service. There are price models that best match to only one type of service and others that could be used for nearly every service.

The only price level that can be used unrestrictedly for every service is the flat rate price model, where no volume is being metered and the service is being paid by a periodic transaction. This one is called an unmetered pricing type or flat rate scheme and usually benefits only a small part – actually about 20% – of very active users [15]. These users would also be willing to pay a higher price.

To have a look at the utility of each metered pricing scheme, we will analyze 5 different types of services: Mobile Internet access, voice, e-mail, instant messaging and content. Internet access by a mobile device as already described was first charged by time, then by data volume. But is pricing by volume really the best solution for internet access? Beyond controversy volume-based pricing is the easiest way to charge for mobile data services and it's quite customer-friendly, because no matter if the connection's speed is fast (like with UMTS) or slow (like with GSM), the customer pays only for the data sent and received. But in future – with multimode devices – there will be a mix of datavolume- and location-based pricing, where the customer will have the choice of downloading a file in a more expensive UMTS-net or in the next restaurant with a cheaper local WLAN-connection.

Voice services, such as a phone call to a friend or to a service hotline can also rely on multiple price models. In the last years the price of phone calls decreased to such a low level, that we can now look at (mobile) telephony as a commodity. A little revolution happen, when WLAN will be integrated in mobile phones, because other players like Skype will provide Voice over IP phone Calls for a much cheaper price then the mobile operators. That will lead to further decrease in prices for mobile phone calls. The easiest and cheapest way to charge phone calls is to count the time of the call, which is the way almost every operator does it. Then there are additional possibilities to charge special services like hotlines. In Switzerland phone calls to premium rate numbers with the prefix 0900 and in Germany to 0190 have different (more expensive) rates for service hotlines but also for entertainment such as adult chat service. Furthermore there is the Sunrise myzone-service as described above.

And how does the customer like to pay for e-mail-services? The most accepted price model is data volume based pricing, because all carriers do so. But who will be able to understand a price based on 10 kilobyte packets? [15] A better solution would be value based pricing, so that maybe one e-mail without attachment costs a specific price and one with attachment costs a little bit more. This solution lets the user overview his/her costs better.

Another fast diffusing application is instant messaging. The possible killer-application of 3G can best be measured by data volume and costs can be easily accumulated to internet access. But here also value based – per message – pricing is a possible option. Content in the sense of news, pictures, games, ringtones etc. etc. is best being measured by the value it has for the customers. Although downloading a game requires much more bandwidth than viewing news-headlines customers' willingness to pay will be much less for a game and much more for news, than the huge difference in data volume.

7.3.8 Payment

Topic of this chapter is how the charges are being collected by the service providers and which intermediaries play essential roles in these transactions. Billing systems are at the core of operator processes and, therefore, must be stable, carrier-grade and operate in heterogeneous environments too [15]. Therefore some alliances between mobile operators, content providers, banks and other partners emerged to standardize the payment process.

To further analyze payment solutions we have to gain insight into mobile payment scenarios: First, the mobile commerce scenario: We can pay **mobile applications and services** (e.g. games) over the air. Second, the electronic commerce scenario: We can pay **goods and content** (e.g. CDs) by credit card or by e-payment, instead of paying cash. Third, the stationary dealer scenario: We can pay with our mobile device directly **in the store or in front of a vending machine** instead of cash-, cash-card- or credit-card-payment. And finally the C2C Scenario: We can **pay each other, borrow money from someone or lend money** to someone [17]. These scenarios help us to classify the most important m-payment segments, which are: the premium segment, with payment solutions for games, mp3-downloads, etc., the remote segment for CDs, books, clothes and other mail-order goods and the point-of-sales (POS) segment for payment solutions directly in taxis and grocery stores. Payment Methods for these segments are as different as the segments themselves and are provided by different providers. We can distinguish between prepaid, postpaid, credit card and direct debit payment. Prepaid solutions are one way for independent stores to sell goods of low value. Such a service could be downloading ringtones, games and background-images, like Jamba (www.jamba.ch/.de) supplies it, thus it best matches for the premium segment and is handled by the service provider or the mobile operator. Postpaid payment is a possibility for the remote segment, where the supplier already knows the consumer or where contracts exist, which assure the supplier. If the mobile operator intermediates, this transaction process could be made safer, because the operator has more information about its users. But not only the remote segment can profit from postpaid m-payment solutions, the POS-segment will also get its revenues thru this method. Most frequently used for electronic transactions is payment by credit card. This is the best solution for high amounts and also useful for the remote segment or the POS-segment. Direct debit is more a local than a global method of transaction, but if the mobile operators and banks of one country play together, this would be a very powerful solution for the remote, the POS and probably also the premium segment.

Now we have some insights what solutions could be used for which transaction. But who will provide those solutions in future? Here also, mobile network operators are in a strong position, because it's the easiest way for the customer to charge the telephone bill. There's just one negative point for the operator: it is difficult and expensive for an individual operator to implement such a payment mechanism, so the better solution is to find some partners and implement it together or create a joint venture as payment provider. If the operators do not cooperate, other external payment providers will take over this place.

Recently different players are working on payment solutions or have already developed a payment structure. Some banks come with their own solution to anticipate the mobile operators' intention to control the market of mobile transactions. VISA is experimenting

together with Barclays Banks and others on a VISA Cash-Smartcard. Palm has been testing m-commerce payments using the Palm PDA as the holder of digital wallet information. Such information can be beamed over the IR port to merchants with Palm-compatible terminals for payment. Nokia equips its phones with a digital wallet too. Browsing through the internet, the user can buy a concert ticket using the VISA-information stored in the wallet, then download the ticket directly to the wallet and at the day of the concert walk through the turnstile while RFID-technology identifies the user and his ticket contactless. There are many open questions about how transactions will be processed and the main question is still in the room: Who will dominate the handling of payments?

Now, having the information about what will happen in the near future, we will venture an outlook into the 4th generation of mobile technology. We will do this in the next chapter by introducing three scenarios about what could happen in future.

7.4 Forecast into 4th Generation

7.4.1 The 4G Issues

Before we try to examine all of the emerging 4G technological and social issues, let us first take a short glimpse on to the existing 3G infrastructure.

Deployment of International Mobile Telephony 2000 standards for third-generation wireless networks may begin these years, with NTT DoCoMo planning to have a nationwide 3G network in Japan. Third-generation networks offer multimedia transmission, global roaming across a cellular or other single type of wireless network, and bit rates ranging from 384 Kbps to several Mbps. The 3G infrastructure is as we see just about to be widely deployed. Having said that, we see, that there already exists an infrastructure that matches most of the business and private needs for wireless communication.

But meanwhile, researchers and vendors are expressing a growing interest in 4G wireless networks that support global roaming across multiple wireless and mobile networks-for example, from a cellular network to a satellite-based network to a high-bandwidth wireless LAN. With this feature, users will have access to different services, increased coverage, the convenience of a single device, one bill with reduced total access cost, and more reliable wireless access even with the failure or loss of one or more networks. 4G networks will also feature IP interoperability for seamless mobile Internet access and bit rates of 50 Mbps or more.

Problems

When we mention 4G wireless communication, we have to be aware of the fact that such kind of infrastructure like the 4G wireless infrastructure will not be deployed until 2006 and maybe even later. Till that time significant effort in research and development has to be made in order to be able to deploy that kind of infrastructure and to enable users, both business and private, to communicate and benefit from 4G wireless communication.

By looking at the issues and problems of the current 2.5G and 3G infrastructures already deployed and in use, we can derive some of the possible but important problems concerning 4G wireless infrastructures and thus issues resolving multiple heterogenous wireless networks:
access, handoff, location coordination, resource coordination to add new users, support for multicasting, wireless security and indentification, network failure and backup, pricing and billing and certainly one of the most challenging problems facing deployment of 4G technology: how to access several different mobile and wireless networs.

By looking at these problems concerning 4G wireless infrastructures, it becomes obvious, that new technologies and devices have to be developed. But by looking further at these problems, it also seems, that whole new network architectures have to be deployed in order to implement the features required to solve the problems mentioned above. So, network architectures will play a key role. Having said that, lets look at three possible network architetures, that might exist inthe future.

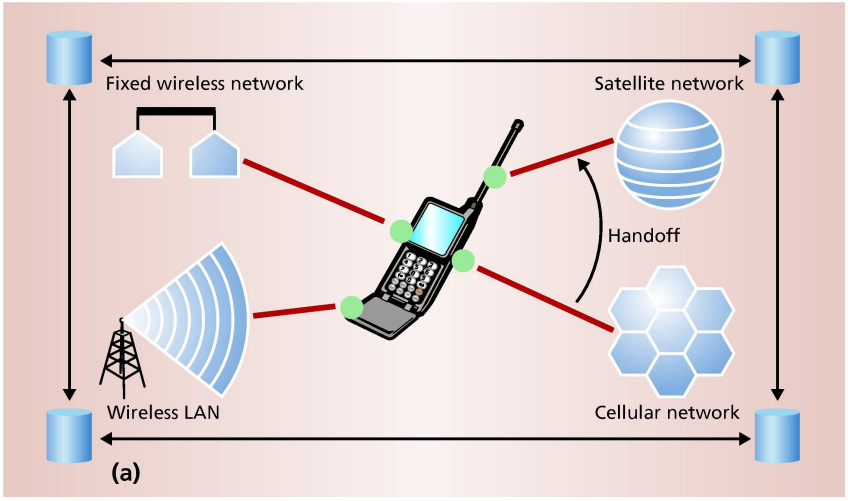

Figure 7.1: Multimode devices

The first possible architecture is one using a *multimode device* (Abbildung 7.1). One configuration uses a single physical terminal with multiple interfaces to access services on different wireless networks. Early examples of this architecture include the existing Advanced Mobile Phone System/Code Division Multiple Access dual-function cell phone, Iridium's dualfunction satellite-cell phone, and the emerging Global System for Mobile telecommunications/ Digital Enhanced Cordless Terminal dual-mode cordless phone. The device itself incorporates most of the additional complexity without requiring wireless network modification or employing interworking devices. The good thing about this architecture is that it may improve call completion and expand effective coverage area. It should also provide reliable wireless coverage in case of network, link, or switch failure.

The second possible architecture represents an *overlay network* (Abbildung 7.2). In this architecture, a user accesses an overlay network consisting of several universal access points (a UAP performs protocol and frequency translation, content adaptation, and QoS negotiation-renegotiation on behalf of users). These UAPs in turn select a wireless network based on availability, QoS specifications, and userdefined choices. The overlay network performs handoffs while the user is moving from one UAP to another.

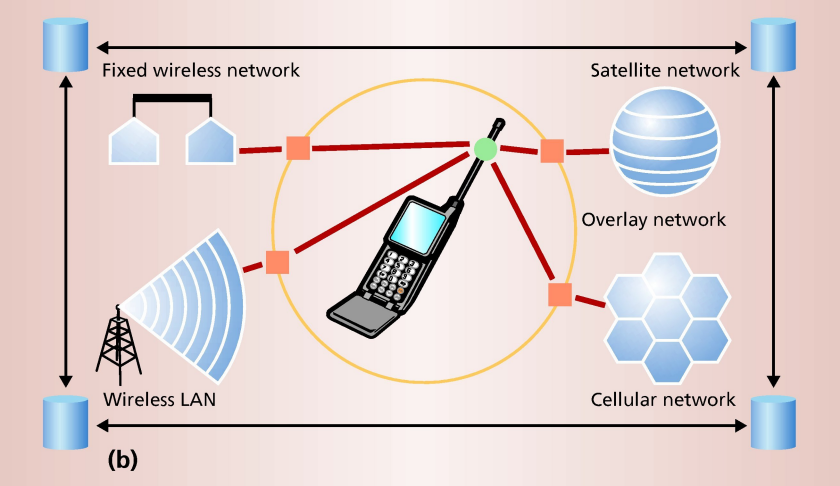

Figure 7.2: Overlay network

Finally, the third architecture uses a *common access proocol* (Abbildung 7.3). This protocol becomes viable if wireless networks can support one or two standard access protocols. One possible solution, which will require interworking between different networks, uses wireless asynchronous transfer mode. To implement wireless ATM, every wireless network must allow transmission of ATM cells with additional headers or wireless ATM cells requiring changes in the wireless networks. One or more types of satellite-based networks might use one protocol while one or more terrestrial wireless networks use another protocol.

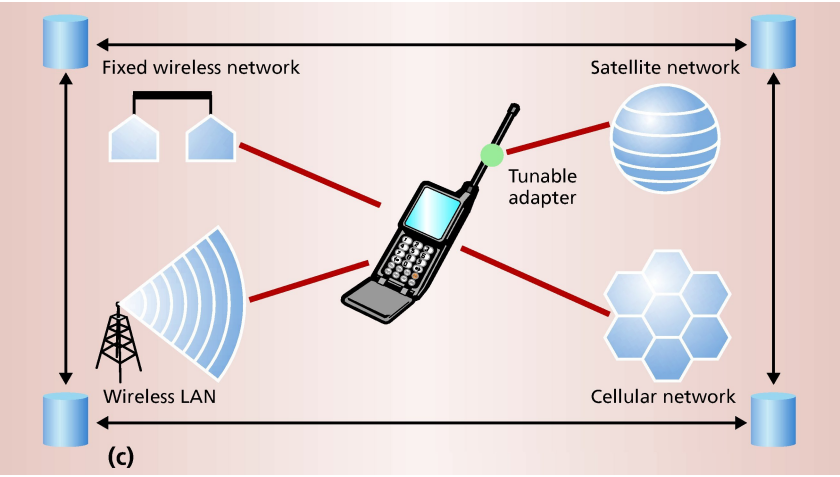

Figure 7.3: Common access protocol

Open Questions

So far we have seen possibilities for implementing 4G network architectures and 4G technology. Yet several open questions have to be answered. We will not try to give an anser to these questions in this paper, though. The aim of this subsection is to provide the

interested reader with an awareness of the problems that may arise during development and implementation of 4G technology.

One of the major challenges will be the support of *Quality of Service (QoS)* due to varying bit rates, channel characteristics, bandwidth allocation, fault-tolerance levels, and handoff support among heterogeneous wireless networks. QoS support can occur at different levels: packet, transaction, circuit or user level. Ergo, the open question here is: How to ensure QoS at different levels among heterogenous wireless networks? Another open question is the one concerning the *handoff delay.*The delay can be problematic in internetwork handoffs because of authentication procedures that require message exchange, multiple-database accesses, and negotiation- renegotiation due to a significant difference between needed and available QoS.

Other questions also arise. Questions such as how to design the new 4G wireless devices to function properly and to be easy to use. What would these devices look like? What new technologies have to be developed? Will the new 4G devices be something competely different or will they be designed just as the existing 3G devices only capable of handling more bandwidth?

7.4.2 The Scenarios

So far we have examined the possibilities and problems looking from the present into the future. What we will do now, is to travel into the future and examine what might happen in a world where 4G technology is already deployed. But we have to be carefull by doing so. One reason why, is because it is verry difficult to identify reasonable assumptions for user behaviour and telecommunication systems, say 8 years from now, that can themselves identify relevant research topics.

Scenarios are thus a tool to explore plausible possible futures, by identifying key technical and social developments that are needed for them to be realized. The point of scenarios is not to predict the future but to create an awareness of what future developments may be possible. The focus however in this paper is to show some possible new opportunities to B2B and B2C communication in the future but also to point out some negative outcomes.

Introduction

The purpose of this scenario set is to describe the possible environments to the computing and communication systems at the time the research outcome may come to market,lets say in 2010. These three scenarios each cover three perspectices:

First, an overview of the whole technological, political and economical system in 2010 is given. This offers a bird's eye view of technological systems, regulations and general development. Second, a scene in telecommunication business is depicted.

Here the aim is to show what kind of issues that are relevant to the business actors. Third, the life of an ordinary citizen is described. The idea of a scenaro is tryed to be shown in

a more natural way. This will allow the interested reader to shape her understanding of these scenarios in relation to her interests and needs. Each scenario depicts a different future. Yet similarities may occure between them. And of course, all parts of a scenario are presented as though 2010 had already arrived, so the interested rader should not confuse the happenings with the present market and technology circumstances.

The three scenarios would be: Anything goes, Big Brother protects u from little brothers and Pocket computing. Lets look at each scenario a little closer.

7.4.3 Anything Goes

Nowdays, in the year 2010, the products have a strong market power. The hardware is generic, so the functions and services are in software. There is a high development pace, a large amount of de- facto standards and a transparent access to the network.

Society and technological background

Over the last decade, there has been a strong development in software solutions for wireless communication equipment. Due to this, the hardware used in transceivers is almost generic and the differences between different networks and systems lie mainly in the software. This has led to a technological system where most problems are resolved through software adaptations instead of tailored hardware. Besides a number of large (regional/national) operators new actors have turned up. Small niche operators offering fairly local wireless networks, aiming at groups of small firms or housing estates, have turned up. From that fact on, many consumers decided to use miniature wireless LANs to cover their homes or their workplaces at home. Almost everybody in the industrialised countries regularly uses information and communication services either through fixed network access or wireless access. This development has increased the demand for transparent access to services regardless of what access system is used. The net has become a very important forum for marketing, education, entertainment, etc. since people of all ages and from all social classes use it.

Inside telecommunication business

Here we show a scene from a meetig room in the European office of a large telecommunicationequipment manufacturer. Well, recent technological progress has made it possible to develop a new modulation technique that allows an increased capacity in multimedia cellular networks by approximately 20%. The representative of a small telecommunication operator is ofcourse interested in this new technology. The discussion moves into questions regarding how the telecomm operator can implement the new solution in the network. It turns out that the Hardware alredy deployed is capable of handling the new software.

The main topic here is the new software and its implementation. The hardware however is of little concern as it is capable of handling the new software. The point here is

that development and implementation of 4G wireless technologies has made the business requirements shift from hardware development to software development and implementation. All B2B handoffs and transactions are handeled trough 4G software. Opportunities arise, when looking at the decreased implementation costs and time when moving to a "'new technology"'. Finally, an agreement was found and the new software was bought.

Everyday Life

This perspective of the scenario shows a normal working day of Mr. Anders who installed his personal WLAN home kit fiew years ago. Its 7:00am on a sunny morning. Mr. Andes just woke up. After a short glimpse into his room he started searching for his glasses. Anders finally finds his glasses. However, they have nothing to do with any vision problem; he never had one. Using Virtual Vision glasses is just the easiest way to know what's up today. "'Oh no, it's 7 o'clock! RADIO shut up!"', he shouts. Instantaneously the music stops. Anders is awake by now and decides to check his mail while still in bed. "'CHECK MY MAIL now that I am awake"', he says. Anders was getting ready to move his hands using the virtual mouse that he could see in front of him through the glasses, when at last his list of e-mails, voice mails and letters appeared in front of him. Lets spend some word on those Virtual Vision glasses. Well, eight years ago (2002) the use of this kind of glasses was verry limited.They just had a transparent screen that used a mirror effect. But today's glasses are much more efficient and broader in use: an opaque and transparent virtual image, a virtual desk and keyboard, integrated microphone and earphones using speech recognition. While looking trough his mailbox he finds, among lots of spam, a message from the manufacturer of his VV glasses telling him, that a new update is ready to be installed. He´s being asked whether to install the update or not. "'Yes, go ahead."' Mr. Aders grumbles.

Suddenly, everything goes black. Something in the upgrade was incompatible with the rest of the software. It's not the first time it happens, and Anders swears loudly. He takes off his glasses, gets out of bed and walks to the computer in the corner of the bedroom. After fiddling around a little, he manages to get his glasses back to a fairly stable state.

After Anders got out of bed and finished checking hi mailthe home appliance agent reminds Anders that his coffee has been ready in the kitchen for a while. On the way towards kitchen, in the corridor, his home network agent appears automatically to advice him of the corridor antenna malfunction, which it has done every other day lately. Anders starts complaining about that antenna asking when the would finally build something cheap and reliable. His shopping search agent appears a few seconds later. Anders reads "'Best price at the moment: HOMETEL 155Mb/s V2002 compliant. 10 Euro. Available in stock. Delivery: 1 day. Buy?"'. "'OK, buy it and take the money from my Central Europe bank account."' The simple everyday words were understood perfectly and he is asked for one more confirmation.

After work Mr. Anders arranges a virtual dinner with his girlfriend. It is a very convenient way of keeping in touch with eachother, trough virtual meetings, but it is however no match for the "'ral thing"'. She is in New York, and he in Stockholm, so while she is having breakfast, Anders is having dinner.

Here we se how 4G wireless technology can benefit private users by offering rich media content that is accessible easily and reliable. For all communication, both business and private, mail, telephone, voice messaging, instant messaging and so on only one 4G device (VV glasses) is needed. But also the intelligent home agents, that tell you every information about your household as soon as you SAY something, make life nowdays in 2010 more convenient.

One thing however remains to be suspicious off; what about the penetration of the private sphere and personal information by linking his entire household and his virtual office at home to the net? A question we'll get back to at the end of this paper.

Deriving from trends

As mentioned before, all scenarios are based on different trends that keep going on. Lets look at some of the trenad underlying this paricular scenario.

Globalisation is one of the underlying trends. Here globalisation manifests itself trough efforts to establish word wide communication, mostly wireless. Companies sell their products globaly and thus are competing on international markets. The main product they are selling is clearly software as hardware is generic and capable of handling all sorts of software.

The second underlying trend is standardization. As schown, companies compete internationally. This implies a bit the existance of big companies with lots of marketshare and influence. Standardization moves in the direchtion of de- facto standards established by those big players among the word wide companies.

A third trend is the development of intelligent household appliances, that can communicate with eachother. All household appliances, such as refrigurators , can communicate with the internet an tell you if they need repair or maintenance or if the milk in the fridge is ready to be thrown away.

7.4.4 Big Brother

Essence of scenario

In this scenario we have a different world than in the one before. Here, the market is limited and controllen by political authorities. Personal integrity becomes no.1 priority in all systems. There is global security, performed international governmental organisations. The development rate is low and there are very fiev operators in the market.

Society and technological development

The development towards open information society looked great,19 but suddenly a couple of years after the millennium turn, it suddenly went to far and the security of the individual

was threatened. By this time it was easy to find almost any information about a person or a company either directly via the WWW, but easier and more efficient by buying the information from one or a couple of the information traders. These were companies specialized in collecting information and selling it. Information were found by scanning email, checking surfing habits and even by mapping behavior when playing the popular net games. There had also been some major cracker break-ins proving that "safe" sytems were not that safe.

For these reasons authorities had to act. By 2009 the governments had strong legislative control over the communications resources such as the available frequency bands and encryption. The new society were built on a new Information Constitution connected to the new Security and Integrity standard ISO 99000, built on previous quality standards. The new system was synchronised with most developed countries.

Inside telecommunication business

In order for the government to be able to control the telecommunication operators very few were certified. So here we show again a meetingroom in some Stockholm office. In the northern parts of Stockholm construction is going on. A new suburb is planned and there will be new offices, apartment buildings and shopping centres there. Since 2007 the demand for mobile telecommunication has been stable, but in this case there is a need to expand the network capacity in that particular area. The representative of the largest telecommunication operator in Sweden is here to buy new base stations for his network in order to expand it into the new suburb.

The main topic in this conversation focuses on new base stations and their implementation into the exsisting network architecture of the operator. Here, software modifications are and their implementation into existing devices is of no greater concern, as far as UMTS is the standard that has prevailed and is capable of handling nearly all neds of nowdays wireless communication. Another topic were the programming interfaces for the new base stations. But since the standards haven´t been changed, this matter is solved.

Everyday Life

Again we introduce our Mr. Anders. It is a normal Wednesday morning in May, 2010. Anders is woken up by the alarm clock that turns on usually at 7.00 am. Slowly he gets out of bed, and heads for the shower. After the shower, almost fully awake, he heads for the kitchen. He grabs what look like a calculator, and type in some number... 'OK, 146539 is my ID-card PIN today' (Used to activate the ID-card) After a while a free copy of the big national newspaper is printed on his printer. He only prints the parts that he is interested in, but he receives everything in order to not reveal his personal interests. He flips through it while eating his breakfast.

After brakfast and newspaperreading, it was time to go to work and anders is heading for the subway. At the entrance, he takes out his ID-card and the gate recognises the period ticket stored in the card and lets him though. During the last years it has been popular to

store different functions in the ID-card since they are able to provide secure transactions without revealing the identity of the user. Finally getting to the office building. At the entrance he again sweeps his ID-card and enters his pin code to get in the building. Security checks are everywhere, Just to get from the front of the building to his office, he needs to sweep his card five times. A new government technology was announced a few weeks ago: Smart badges will be available soon.

Sitting in his office, Anders logs onto the network. He is notified that there are a few new email messages waiting for him. There is one message where he is asked to reply and contribute with comments. He spends a couple of hours to write down the reply. When he is finished he pushes the "'send"' button and a large box shows up on the screen. It informs him that he should not pass any personal information in an email. It also informs him that he should be aware that sensitive company information should not be sent in emails and that all external email will be screened for sensitive data. He knows that the email will be checked by the internal information control branch, which co-operate with the government.

Troughout the working day of Mr. Anders we come over a fiew life circumstances, typical nowdays: Personal data can only be read by the local authorities and only if the particular ID card is being used. Shopping in the grocery store, buying tickets for the cinema and other transactions are made solely with the personal ID card there are no spam mails in the mailbox, no unsecure or obscure internet sites, even the pay-per-view television was outlawed in 2007 because they could get too much personal information.

Deriving from trends

One of the constraining trends of this scenario is the increased use of information trading. The first impression when we look at the informatio trading mechanism is quite positive. Needed information can be gathered on the net. Everywhere. All sorts of information about everything and averyone can be collected from the net. This seems good, as though we can enrich our personal knowledgebase or firms can establish a good market knowledge trough fast, rich and easy to assecc information. Yet caution is adviseable. Especially at the point, where information trading begins to degenerate and where, as the direct recult of degeneration, sensitive information becomes available. Beyond that point access to sensitive information becomes almost to easy.

The second trend driving this scenario is globalisation. Here globalisation is decreased as authorities have to limit cross-border communication Standardisation is increased as international organisations and governments enforce a new integrity and encryption policy. Firms split off into smaller regional units, which in turn cooperate with the local authorities. Products can hardly be sold worldwide.

What we see here, is that 4G technology is being restricted and used in a very narrow way. Although we have 4G wireless infrastructures, they are being used only in spezific ways. So this scenario shows how overall and worldwide roaming and communication (one of the goals of 4G technology) can lead to almost total restriction and security barriers as communication becomes to risky for ones private sphere and pesonal data integriy.

7.4.5 Pocket Computing

Essence of scenario

In our last scenario, the market is well developed and controlled by big operators. The situation on the market is as follows: bandwidth is still expensive and there are highly differentiated service and pricing levels. Specialised service providers also provide equipment for specialised purpose.

Society and technological development

Third generation systems are mature, IMT-2000 has been reviewed several times since its commercial start in 2002, still delivering wireless up to 10 Mbps in offices, 2 Mbps in cities, and 384 kbps in the country side. However, the universal mobile idea has never reached the expectations. A large variety of different mobile solutions are available. GSM is still widely used with some new features but serious lack of bandwidth. Hence, it remains inappropriate for mobile multimedia needs. The different wireless networks compliment each other, but multi-mode solution is relatively expensive. Common people prefer to pay for reduced services at minimum price.

Because the cost of antenna sites has risen enormously, telecommunication operators invest more in services than in infrastructure. Therefore, to match to the offer, the price of the broadband access has been risen. Only businesses want to afford all-in-one devices and pending services. Most of the people prefer to have light and small terminals to carry everywhere. Ad-hoc networking capabilities make it easy to wirelessly connect camera, display, keyboard or other computers within a short range. There is also a growth of hidden communication, i.e. cars communicate with each other, household appliances etc. Those communications turned out to be very helpful; they increase the security and are often partly financed by authorities. These devices are often wireless. They use relatively low bit-rate communication and simple protocols.

Most companies are global players and are present in every country with marketing service and studies. Products are marketed worldwide and global corporations have invested in worldwide corporate networks to coordinate the information and limit information leakage.

Inside telecommunication business

During the past few years, most of the big corporation tried to grow by acquiring little businesses around the world to gain locality. Competition is extremely intensive in the computer and telecommunication industry. The number of big companies in this sector is limited. Small companies are competing locally and end up by being bought.

Again we have our meeting room where two big companies discuss an eventual merger. One of the two companies is a huge Asian computer/router/etc. Company and the other is a European data warehousing company. The goals are to grow both horizontally, and

vertically. They want to compete with the telecommunication establishment by increasing the WLAN network use and extend the idea to a non-cellular mobile phone/data network. The advantage is that the WLAN technology is adaptive and therefore needs very little network management. Another reason for the merging is that the two corporate communication networks are complementary. The lack of fixed lines in Asia could be provided by merging the networks and vice versa.

But also other concequences are expected, such as: Devices talk without problems to each other, the next step is to extend it to a worldwide phone/data net. Ad-hoc networking will give a good coverage solution combined with the available joint infrastructure. Also 24h around-the-world software development may become possible.

Every day life

Here we show two two different kinds of people to be able to illustrate in more detail the life circumstances nowdays. We make this distinction of people- profiles in order to make the different market offerings and available services more visible and to enable the interested reader to distinguish between them.

First there is our educated and well paid consultant in investment banking and secondly our engineer with academic background but working as an employee for a telephone operator. The first one we call the Businessman the latter the New MiddleClass Stanislaw, our Businessman is 45 years old, single, and resident in Luxembourg at the moment. He speaks a few languages but mainly communicates in English. He has World-wide clients. He is moving from project to project and country to country.

The thin notebook is his desk always virtually connected to the network. He uses a multimode phone/modem that allows him to phone from every place and connect most of the time with a correct data bit rate to the Internet and to the central server (or its mirror). His company pays for the services. He needs fast access to information. Global communication and traveling are vital for his business. He can afford secure and reliable communication for premium rates. He expects that global products can be found everywhere he goes without local adaptation.

Aldous is electrical engineer and works for a telephone operator in the US. He belongs to the so-called "'new middle class"', accounting for approx. 50% of the average US-European population. Aldous has a reasonably high income to afford vacation travelling, hobbies and the little nice things in life, such as a car, a cell phone, cinema, theatre, and lots of friends. He also has a 11 year old son (Thomas). Thomas is of the new communication generation. He is young and therefore a good target for the network game industry. Thomas couldn't imagine any game that cannot interact with his friends. It is not funny to play alone. The devices Thomas and his friends are playing with are using direct LOS connections and have to be played in a room to avoid interference with other eventual players.

In the past fiev years, more and more new terminals are available, each offering different service and quality level. Some of them have software update capabilities and give

the possibility to extend the original service. They are expensive as the same terminal competes with different services, but it is handy to have only one device in the pocket.

Aldous has a basic Mobile Phone (GSM and DECT compliant) and could access high bit rate WCDMA services but he avoids it because it is expensive. He has a fixed network connection at home and at work and he thinks it is enough. There is, however, one thing Aldous has problems with: whenever he accesses the network, he is traced by hyper-companies, it is part of the cheap service contract. He knows that if he buys something he will receive mail or post the next days advertising for accessories, or better deals from concurrent. If he could afford his secure personal connection, he would go for it at once. But not only the mobile phone services are attracting Aldous. He even connected a daily kitchen forum directly to the shopping agent so he can use the interactive TV show to pepare the dinners.

The main point in this perspective of the scenario is, that some service providers are enabled to charge premium prices for a fast and reliable access to the net while others compete in delivering cheap services but profit from additional advertising a user paying low prices has to live with.

Deriving from trends

The Pocket Computing shares most of the trend development with the Anything Goes scenario, with the difference that the development does not reach all people. A word on Standardisation: Standardisation is ruled by big operators. So this means that we have de facto standards. The development of communicating Appliances is slower because the technology development trend lies clearly in developing software and services rather than appliances that can communicate with eachother.

Hopefully, the scenarios were fun to read but also interesting and insightfull. However, they show three possible but different futures that, derived from trends in technology development, societal structures and economical changes, are quite possible and real.

7.4.6 The Outlook

After looking into the futue of wireless communication and showing new possible opportunities for both B2B and B2C communication let us return to the present. In this section it will be tryed to give a carefull and subjective outlook into the future of wireless communication.

We start our outlook by first looking at the possible technological development and here briefly adress the open questions related to it. Then we compare the 4G and 3G architectures shortly and finally conclude this section.

Technological Development

After examining possible technologies that might come to use in the scenarios and looking at present progress in developing both wireless communication devices and software (protocols, etc.) a subjective oppinion is formulated: 4G wireless technology will be based mostly on software. Devices, such as Mobile Phones, PDA's and Notebooks, will not differ much from todays devices; perhaps they will be different in design an use. The main difference however will be the software used, the interfaces provided and the content shown. Wireless access will be posible with all kind of devices, anywhere and anytime. As said before, software is going to be the main carrier of functionality and content. New security standards will be made to make wireless communication more secure and reliable.

Open questions

However, there still remains a number of open questions concerning 4G technology. In earlier sections of this paper we have adressed some of them such as the QoS support or Handoff delay when switching among different networks.

Other question arise while looking at the technological development, especially in cration of communication protocolls and security standards. What if a security standard will be "'just"' a de facto standard posed by some fiew big operators? How will the WEP , i.e. Wireless network security leaks be filled? What will be the programming interfaces for software on recievers or end user devices? How will the media content creation and packaging value chain beaffected by new devices or software? There are clarly more open questions that need answering but this would leave the scope of this paper.

	3G (including 2.5G)	4G
Major Requirement Driving Architecture	Voice driven; data was always add-on	Converged data and voice over IP
Network Architecture	Wide Area Cell- Based	Hybrid - Integration of Wireless LAN (WiFi, Bluetooth) and wide area
Speeds	384 Kbps to 2 Kbps	20 to 100 Mbps in mobile mode
Frequency Band	1800- 2400 Mhz	Higher frequency bands (2-8 GHz)
Bandwidth	5- 20 Mhz	100 MHz (or more)
Switching Design Basis	Circuit and Packet	All digital with packetized voice
Access Technologies	W- CDMA; 1xRTT; Edge	OFDM and MC-CDMA (Multi Carrier CDMA)
Forward Error Connection	Convolutional Rate 1/2, 1/3	Concatenated coding scheme
Component Design	Optimized antenna design, multi- band adapters	Smarter Antennas, software multiband and wideband radios
IP	Air link protocols incl.IPv5.0	All IP (IP6.0)

Figure 7.4: Technical comparison of the 4G and 3G Technology

Current infrastructure vs. 4G infrastructure

Here we show a table (Abbildung 7.4) that depicts a comparison of 3G and 4G technology, hereby focusing on more technical data such as speed, bandwidth and so on. The goal

here is to give a direct confrontation of the two technologies in order to better distinguish between them and to show what is possible with 4G technology.

Conclusion

Ultimately, there can be no differentiated and precise assertion on what the future brings in wireless communication. But we can make pretty good assumptions, based on facts and happenings, we see nowdays. We saw possible futures and we saw, how people might live by using and interacting with all those new technologies and devices.

However, there is much more to say in this mather, but the focus in this paper was to provide an insightfull overview of possible happenings and to give the interested reader the awarenes of what the future in communication might become. Of course interested reader is nevertheless encouraged to further investigate this topic.

Bibliography

[1] Wikipedia, Online Encyclopedia, www.wikipedia.org, 2004

[2] Jörg Roth: Mobile Computing, Grundlagen, Technik, Konzepte, dpunkt.verlag, Heidelberg 2002.

[3] Pierre Lescuyer: UMTS, Grundlagen, Architektur und Standard, dpunkt.verlag, Heidelberg 2002.

[4] Unisys, Annual Report 2000 Glossary, www.unisys.com/annual/annual2000/glossary/, 27.12.2004

[5] Indbazaar, Netguide, Glossary,www.indbazaar.com/netguide/index1.asp?catid=12 &alfa=M, 27.12.2004

[6] Mc Graw Hill, Education, highered.mcgraw-hill.com/sites/0072464011/student_view0/ chapter4/glossary.html, 27.12.2004.

[7] Nokia, Support Glossary, www.nokia.com/support/glossaryCDAMaster/0,„lang_id= 46&keyword=mobile _commerce,00.html#mobile_commerce, 27.12.2004.

[8] Mobile Commerce I, Sommersemester 2004, Lehrstuhl für M-Commerce & mehrseitige Sicherheit, Prof. Dr. Kai Rannenberg, http://www.m-lehrstuhl.de, Johan Wolfgang Goethe-Universität Frankfurt am Main.

[9] Siemens Program System Engineering, Intelligent Net Working, http://www.pse.siemens.at, 27.12.2004

[10] Nicholas D. Evans, The M-Business Evolution, August 2002, http://www.developer.com/ws/other/article.php/10949_1446771_1

[11] NTT DoCoMo, Japan's premier mobile communications company, http://www.nttdocomo.com, Core Business, i-mode

[12] TagesAnzeiger, Newsticker, 9.12.04 17:21, www.tagesanzeiger.ch

[13] Stuart J. Barnes, The mobile commerce value chain: analysis and future developments, International Journal of Information Management, Volume 22, Issue 2, April 2002, Pages 91-108

[14] Falk Müller-Veerse, Mobile Commerce Report, Durlacher Research Ltd, Press Release 1999, www.durlacher.com

[15] Dr Erkko Autio et al., UMTS Report, Durlacher Research Ltd, Press Release 2001, www.durlacher.com

[16] Juniper Research Ltd, Mobile Commerce & Micropayment Strategies, Publication Date: August 2004, ISBN: 1-904405-36-3, www.juniperresearch.com

[17] Khodawandi, D.; Pousttchi, K.; Wiedemann, D. G.: Akzeptanz mobiler Bezahlverfahren in Deutschland; in : Proceedings zum 3. Workshop Mobile Commerce; Augsburg 2003

[18] www.s3.kth.se/radio/4GW/public/Papers/ScenarioReport.pdf Maxime Flement, Fredrik Lagergren, Rickard Stridh, Olav Queseth, Matthias Unbehaun, Jiang Wu, Jens Zander: Personal Computing and Communication

[19] Eurotechnology Japan K. K.; http://www.eurotechnology.com/4G/

[20] www.ee.oulu.fi/˜ skidi/teaching/mobile_and_ubiquitous_multimedia_2002/issues_in_emerging_4G_wireless_networks.pdf

[21] 4Gcouk Limited; http://www.4g.co.uk/

Kapitel 8

Gruppenstrategien für Online-Business am Beispiel Mobilfunk

Robin Bucciarelli, Luzius Hausammann, Manuel Donner

In dieser Arbeit werden in einem ersten Schritt die Grundlagen von strategischen Allianzen beleuchtet. Dabei werden die wichtigsten Aspekte, welche zu einer Allianzbildung führen, kurz erläutert. Eine wichtige Rolle dabei spielt die Unterscheidung zwischen der ökonomischen und der juristischen Sichtweise. Die dabei gewonnenen Erkenntnisse werden unter wettbewerbspolitischen Aspekten analysiert. In einem zweiten Schritt wird das theoretische Wissen auf den Telekommunikationsmarkt übertragen. In drei Phasen wird ein historischer Abriss der Marktentwicklung seit der Deregulierung bis zum jetzigen Zeitpunkt dargestellt. Dabei werden die sog. „Global Players" untersucht und deren Allianzverhalten durchleuchtet. Der dritte und letzte Schritt nimmt aktuelle Allianzen in der Mobilfunkbranche unter die Lupe und zeigt deren Struktur. Die „Open Mobile Alliance" bildet den Schwerpunkt dieser Ausführungen.

Inhaltsverzeichnis

8.1 Grundlagen

Seit einiger Zeit lassen sich in der Mobilfunkbranche starke Tendenzen zur Bildung von Allianzen feststellen. Diese ermöglichen eine gemeinsame Entwicklung und Vermarktung von Diensten. Solche Allianzen können die Marktmacht gegenüber Handelspartnern stärken und die Kundenbindung erhöhen. Im Rahmen dieser Seminararbeit werden verschiedene Beispiele für Gruppenstrategien bzw. Allianzen untersucht. Dabei wird versucht zu erklären, weshalb gerade in der Mobilfunkbranche die Allianzbildung einen derart hohen Stellenwert besitzt.

Die ursprüngliche Seminarthematik „Gruppenstrategien für Online-Business am Beispiel Mobilfunk“ wurde auf den gesamten Mobilfunkmarkt ausgedehnt und folglich in „Gruppenstrategien am Beispiel Mobilfunk“ umbenannt. Diese Änderung wurde als notwendig erachtet, da das Begriffspaar Online-Business und Mobilfunk in der Praxis wenige Beispiele liefert.

8.1.1 Einführung und Definitionen

Koalition oder Kooperation im weitesten Sinne steht als Sammelbegriff für jede Form der Zusammenarbeit zwischen Unternehmen zur Erreichung gemeinsamer Ziele. Weinhold definiert Unternehmenskooperation wie folgt: „*Zusammenwirken von selbständigen Gebilden unter Beibehaltung der Globalautonomie und der Aufgabe der Partialautonomie zugunsten des Kooperationsgebildes.*“ [1] Eine etwas engere Umschreibung des Begriffes „Kooperation“ findet sich bei Bischof-Köhler: „*Unter Kooperationen sei eine Verhaltensweise verstanden, bei der zwei oder mehrere Individuen in einer Weise interagieren, die die Wahrscheinlichkeit erhöht, ein gemeinsames Ziel zu erreichen. Es genügt hierbei nicht, dass mehrere Individuen das Gleiche tun, ... , entscheidend ist vielmehr, dass ihre Aktivitäten aufeinander bezogen sind und sich ergänzen.*“ [2] Hierbei wird ersichtlich, dass die Kooperationspartner nicht nur gemeinsame Interessen verfolgen, sondern die Zusammenarbeit auf Freiwilligkeit basiert. Jeglicher Zwang würde unweigerlich zu einem Boykottverhalten des Gezwungenen führen. Kooperationen können demnach als eine freiwillige Zusammenarbeit zwischen zwei Unternehmen zur effizienteren Erreichung von gemeinsamen Zielen verstanden werden.

Der Begriff „Koalition“ charakterisiert dabei die unterschiedlichen Kooperationsbeziehungen wie z.B. Joint Ventures, strategische Allianzen, strategische Partnerschaften usw. Diese sind von den sog. „Mergers and Aquisitions“ den Unternehmensfusionen und Übernahmen, abzugrenzen, da bei Unternehmenszusammenschlüssen die jeweiligen Unternehmen ihre rechtliche Selbständigkeit verlieren und somit als neue (Mergers) bzw. erweiterte (Aquisitions) juristische Personen auftreten. Diese Unabhängigkeit ermöglicht den Unternehmen, in den von der Kooperation unberührten Geschätsbereiche den Wettbewerb aufrecht zu erhalten. Kartelle hingegen entsprechen der Definition einer Kooperation weitestgehend. Der Zweck eines Kartells liegt aber nicht in der Ausnutzung von möglichen Synergieeffekten von Unternehmen, sondern erfüllt primär die Rolle eines Regulator in Form von Wettbewerbsbeschränkungen.

Allianzen sind, wie im vorangegangenen Absatz erläutert, hauptsächlich Kooperationsbeziehungen von rechtlich und wirtschaftlich unabhängigen Grossunternehmen. Dabei kann der Begriff synonym mit Kooperation bzw. Koalition verwendet werden. Der Ausdruck „Allianz“ wurde von Unternehmensberatungsgesellschaften und nicht in der Wissenschaft entwickelt, dementsprechend existiert keine klare Definition. Der Begriff kann für alle möglichen Formen der Zusammenarbeit von Unternehmen verwendet werden. Der Grund für die Entstehung eines solchen Begriffes ist jedoch klar ersichtlich: Die Globalisierung der Märkte und der dadurch erhöhte Wettbewerbsdruck auf die Unternehmen, hervorgerufen durch den Abbau der Handelsschranken bzw. durch die Öffnung der Märkte, führt zu einer Vereinheitlichung des Konsumentenverhaltens. National operierende Unternehmen bekommen zusehends den internationalen Konkurrenzdruck ausländischer Anbieter zu spüren. Auch die rasante Beschleunigung der aktuellen Technologieentwicklungen tragen das ihre dazu bei, dass Ressourcen von einzelnen Unternehmen bei diesem Tempo nicht mehr ausreichen, um im Konkurrenzkampf zu bestehen. Daher suchen immer mehr global tätige Unternehmen nach strategischen Allianzen, um Ressourcen mit gleichgesinnten Unternehmen zu vereinen (sog. Ressourcenpooling).

Eine abstrakte Umschreibung strategischer Allianzen nach Reuter kann folgendermassen formuliert werden: „*Zusammenarbeit von zwei oder mehreren unabhängigen Unternehmen, die darauf abzielt, gemeinsame Wettbewerbsvorteile oder Erfolgspotentiale für die gesamte Geschätsbreite zu erschliessen, oder um langfristig angelegte Bündnisse mit dem Zweck, umfassende Aufgaben für ein strategisches Geschätsfeld gemeinsam zu erfüllen.*“ [3]

Trotz uneinheitlicher Definitionslage können nichtsdestotrotz einige wesentliche und typische Merkmale von strategischen Allianzen festgehalten werden:

- Strategische Allianzen stellen Kooperationen von rechtlich und wirtschaftlich unabhängigen Unternehmen dar. Dabei handelt es sich meist um global agierende Grossunternehmen.
- Die Zusammenarbeit beschränkt sich auf bestimmte Geschätsfelder oder Projekte
- Das Konkurrenzverhältnis bleibt in den übrigen Geschäftsfeldern bestehen
- Unternehmen in strategischen Allianzen sind meist international ausgerichtet
- Die strategischen Zielsetzungen bzw. das gemeinsame Interesse von Allianzpartnern ist von einem Unternehmen allein nicht erreichbar
- Allianzen sind oft zeitlich begrenzt. Dabei kann die Dauer des Lebenszyklus eines Produktes, einer Technologie oder die Zeitspanne eines Projektes das Bestehen der Allianz bestimmen
- Allianzen sind stets vertraglich geregelte Kooperationsformen (siehe juristische Aspekte weiter unten)

8.1.2 Motive für die Bildung strategischer Allianzen

In diesem Unterabschnitt werden die Hauptgründe für die Bildung strategischer Allianzen näher betrachtet. Als Motivation für die Allianzbildung kann dabei die Überzeugung zweier oder mehrerer rechtlich selbständiger Unternehmen angesehen werden, welche angestrebte Unternehmensziele mit einem oder mehreren Partnern gemeinsam besser verwirklichen können als ohne Kooperation. Dadurch stellt diese Form von Kooperation eine echte Alternative zu Fusionen oder Akquisitionen als konzentrative Unternehmensverbindungen dar.([4], S.49) Das Risiko bzw. der Kapitaleinsatz kann somit bei einem schnellen Markt- oder Technologiezugang erheblich gesenkt werden.

Globalisierung

Die Weltwirtschaft ist als ein Markt ohne nationale Grenzen zu betrachten, in welchem ein weltweiter Wettbewerb betrieben wird. International agierende Unternehmen verwandeln sich in sog. „global players“ , welche aus eigener Kraft bzw. mit den in der Unternehmung verfügbaren Ressourcen nicht mehr wettbewerbsfähig sein können. Daher bieten sich strategische Allianzen zur gemeinsamen Ressourcenausnutzung an.

Globalisierung speziell im Informations- und Kommunikationssektor ist die Globalisierung von Unternehmensaktivitäten auch als Reaktion auf weltweit unterschiedliche Deregulierungsstati zu begreifen. ([4], S.51)

Technologischer Wandel

Die durch die Globalisierung neu entstehenden Anbieterstrukturen zeigen, dass durch die Verschmelzung von verschiedenen Branchen und Technologiefeldern eine sog. Konvergenz stattfindet. Diese Konvergenz führt zu einer Absenkung der Markteintrittsbarrieren, ermöglicht branchenfremden Unternehmen leichter in einem neuen Markt Fuss zu fassen und schwächt gleichzeitig die Position etablierter Anbieter.

Technologischer Wandel löst somit strategische Allianzen aus, da Unternehmen gezwungen sind das benötigte Wissen und die Ressourcen aus anderen Technologiefeldern über eine strategische Allianz zu beziehen.

Ausschöpfung von Grössenvorteilen (economies of scale)

Das Hauptmerkmal von Grössenvorteilen ist eine überlineare Produktionsfunktion. D.h. dass economies of scale Kostenersparnisse darstellen, die aufgrund von Größenvorteilen entstehen. Durch eine hohe Produktions- und Verkaufsmenge können ein hoher Marktanteil und die Kostenführerschaft erreicht werden. Die Kostenersparnisse durch Massenproduktion stehen in der Regel denjenigen durch die gemeinsame Nutzung von Ressourcen bei verschiedenen Produkten entgegen. Im Falle der Telekommunikationsbranche entstehen

Skalenerträge bei der gemeinsamen Nutzung einer Netzinfrastruktur, was zu einer Erhöhung der Netzteilnehmerzahl führt und dadurch zu einer Einsparung von Durchschnittskosten. Beide Allianzpartner bringen dabei eine bestimmte Zahl an Netzteilnehmern in die Allianz ein, wodurch die durchschnittlichen Stückkosten, welche wiederum die Kosten für die Unterhaltung der Netzinfrastruktur einschliessen, sinken (Fixkostendegression).

Nutzung von Verbundsvorteilen (economies of scope)

Economies of scope entstehen, wenn zwischen zwei Unternehmen Synergien bestehen, welche ein Projekt oder Produkt zusammen günstiger realisieren können alls alleine.[5] Bestimmte Kosten von zwei verschiedenen Produkten oder Dienstleistungen können durch eine strategische Allianz tiefer liegen als bei der getrennten Erzeugung in zwei Einprodukt-Unternehmen. Dabei spielt die Komplementarität von Ressourcen und Know-how eine zentralle Rolle. Ein Unternehmen, welches zwecks Diversifizierung in einem neuen Markt eintreten will, wird sich zwangsläufig das benötigte Know-how durch die Bildung einer strategischen Allianz erschliessen.

Aufwendungen für Forschung und Entwicklung

Ein Unternehmen, welches im Alleingang Forschung und Entwicklung betreibt, geht sehr hohe Risiken ein, nicht zuletzt wegen den hohen Kosten, welche z.B. bei der Einstellung von hochqualifiziertem Personal anfallen. Durch eine Kooperation im Bereich Forschung und Entwicklung können die entstehenden Kosten und das Risiko auf mehrere Partner verteilt werden und zudem Doppelforschung vermieden werden. Die Fixkosten für den Unterhalt von Forschungseinrichtungen kann somit auf eine grössere Anzahl von Projekten verteilt werden.

Innovationskosten

Hohe Innovationskosten, welche durch die immer kürzer werdende Zeitspanne von Produktlebenszyklen anfallen, müssen durch Erträge amortisiert werden. Oft wird dies durch das sog „appropriabilty-Problem“ erschwert. Das „appropriability-Problem“ beschreibt den Nachteil, den ein Innovator durch einen nicht ausreichenden Patentschutz seiner Innovation hat. Wenn der Zeitraum des Patentschutzes nicht ausreicht, um die anfänglich für die Entwicklung der Innovation entstehenden Kosten zu amortisieren, wird der Investor die Investition nicht tätigen.[6] Hinzu kommt das Problem der Imitatoren, welche zwecks Umgehung des Patentschutzes durch reversed engineering ein Konkurrenzprodukt auseinanderbauen, um die dahinterstehende Technologie nachzuvollziehen und darauf nachzubauen oder sogar zu verbessern. An dieser Stelle helfen strategische Allianzen die Amortisationsrisiken besser unter den Partnern zu verteilen, vor allem wenn sich die Zusammenarbeit auf die gemeinsame Verwertung von Forschungsergebnissen erstreckt.

Konkurrenz unterschiedlicher Standards und Normen

Ein Standard ist eine breit akzeptierte und angewandte Regel oder Norm. Der Begriff findet im allgemeinen Verwendung als Synonym für eine technische Norm und den Bedeutungen Industriestandard und „herstellerspezifischer (proprietärer) Standard“ . Eine Norm ist eine allseits rechtlich anerkannte und durch ein Normungsverfahren beschlossene, allgemeingültige sowie veröffentlichte Regel zur Lösung eines Sachverhaltes.[7]

Bei der Entstehung eines Marktes ist noch nicht abzusehen welche Produkte oder Dienstleistungen sich langfristig durchsetzten werden. Eine schnelle Marktpenetration oder Durchdringung des Marktes durch ein Produkt kann dieses als dominantes Design zu einem de-facto Standard werden lassen. Dadurch werden die Konkurrenzprodukte bereits in der Entstehungsphase ausgeschaltet, was Fehlinvestitionen und Kompatibilitätsprobleme auf Seiten der Produzenten und Konsumenten reduziert.

Strategische Allianzen stellen eine nichthierarchische Möglichkeit dar, um Netzbetreiber, Dienstanbieter, Hardware- und Softwarehersteller in einem fragmentierten, unübersichtlichen Umfeld gemeinsam zur Durchsetzung von Standards zu mobilisieren. ([4], S.56)

Transaktionskosten

Transaktionskosten sind alle Kosten, die in direktem Zusammenhang mit einer Transaktion (z. B. Kauf, Verkauf) von Gütern entstehen. Die Transaktionskostentheorie besagt, dass alle Transaktionen auch Transaktionskosten verursachen. Kosten sind dabei alle Aufwände, nicht nur monetäre. [8] Im Fall der Marktwirtschaft lassen sich Transaktionskosten in die Kosten der Marktbenutzung und in die Koordinierungskosten im Unternehmen einteilen. Dabei sollen obengenannte Kosten den Entscheid über Eigen- oder Fremdbezug bzw. In- oder Outsourcing von Dienstleistungen ermöglichen. Strategische Allianzen bieten dabei eine Zwischenform zwischen Marktbezug und Eigenherstellung bei einer gleichzeitigen Reduktion der Transaktionskosten durch gemeinsame Erfahrung der Allianzpartner an.

Überwindung von Markteintrittsbarrieren

Markteintrittsbarrieren sind Hürden, die Unternehmen den Eintritt in einen bestimmten Markt erschwehren. Der potentielle Anbieter muss Infrastrukturen für die Produktion errichten, ein Vertriebsnetz aufbauen sowie Werbung für die eigenen Produkte betreiben. Etablierte Anbieter müssen sich nicht um diese Hürden kümmern und können daher unter günstigeren Bedingungen ihre Produkte entwickeln. Die obengenannten Hürden eines Markteintritts können durch strategische Allianzen reduziert, wenn nich sogar direkt umgangen werden.

Lernprozesse

Neben den finanziellen und materiellen Vorteilen von strategischen Allianzen sollen hier die Vorteile von Wissenstransfer zwischen Unternehmen angesprochen werden. Da Wissen

ein sehr komplexes Produkt darstellt, welches nicht am Markt gehandelt werden kann bzw. welches sehr schwer mit einem Preis versehen werden kann, ist das Erwerben eines solchen Gutes nur durch Kooperationen in Form einer intensiven Zusammenarbeit möglich.

8.1.3 Erscheinungsformen strategischer Allianzen

Strategische Allianzen können grob in zwei sich ergänzende Ansätze unterteilt werden, dem ökonomischen und dem juristischen Ansatz. Beide Teilbereiche sind zur Bildung einer strategischen Allianz unabdingbar und charakterisieren gleichzeitig ihre Erscheinungsform. In diesem Abschnitt werden obengenannte Erscheinungsformen näher betrachtet.

Ökonomischer Ansatz

Aus ökonomischer Sicht ist es sinnvoll, die Allianzpartner in Bezug auf ihre Position in der Wertkette und auf ihre Branchenzugehörigkeit darzustellen. Dabei treten drei Haupttypen von Allianzen auf: die horizontale, die vertikale und die diagonale Allianz. In der Praxis sind diese Kategorien jedoch nicht immer eindeutig identifizierbar, was Interpretationsspielräume offen lassen kann.

Horizontale Allianzen

Horizontale Allianzen sind Verbindungen zwischen Unternehmen auf derselben Wertkettenstufe. Dabei vereinigen die Allianzpartner ein oder mehrere Funktionsbereiche ihrer Unternehmungen. Je nachdem, welche Funktionsbereiche von der Kooperation tangiert sind, können Forschungs - und Entwicklungsallianzen, Marketing und Vertriebsallianzen, Beschaffungsallianzen und Produktionsallianzen entstehen. Die Bildung einer horizontalen Allianz kann durch eine vertragliche Bindung, eine gegenseitige Kapitalbeteiligung der Allianzpartner oder durch die Gründung eines Gemeinschaftsunternehmens erfolgen.

Abbildung 8.1 zeigt das Beispiel einer horizontalen Allianz mit wechselseitiger Kapitalbeteiligung, wobei beide Allianzpartner auf einer Wertkettenstufe kooperieren.

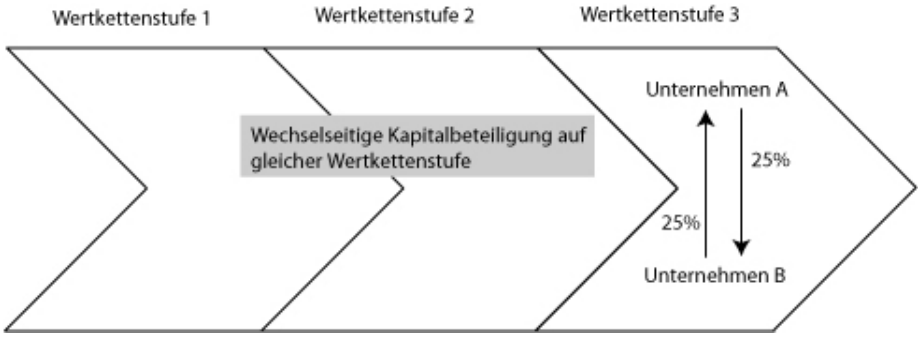

Abbildung 8.1: Wechelseitige Kapitalbeteiligung in horizontalen Allianzen

Abbildung 8.2 veranschaulicht ein horizontales Gemeinschaftsunternehmen, welches die Funktionsbereiche Forschung und Entwicklung vereinigt. Es können jedoch, wie oben erwähnt, beliebige Funktionsbereiche in die Kooperation miteinbezogen werden.

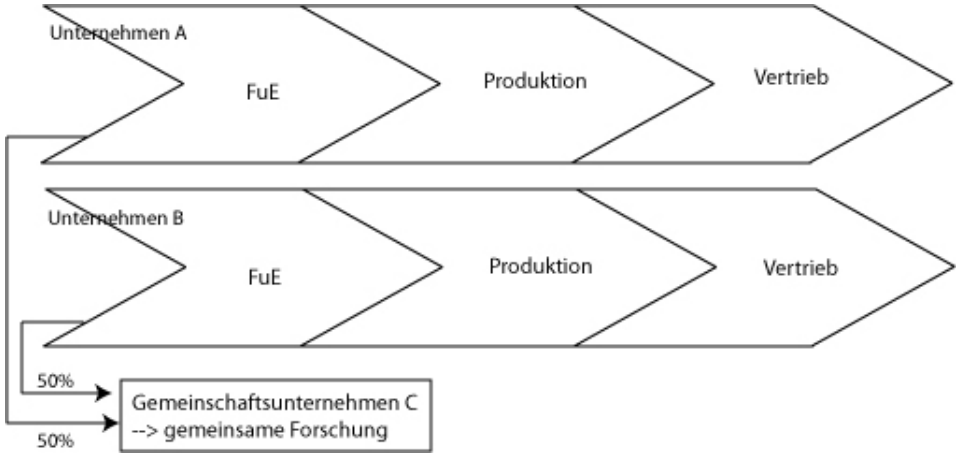

Abbildung 8.2: Horizontales Gemeinschaftsunternehmen

Vertikale Allianzen

Unternehmen, welche auf vor- und nachgelagerten Wertkettenstufen tätig sind und zudem in einer Käufer-Verkäufer Beziehung stehen, können zum Zweck einer partiellen Zusammenarbeit eine vertikale Allianz bilden. Die Bildung einer solchen Allianz kann, wie bei der horizontalen Allianz, durch gegenseitige Minderheitskapitalbeteiligungen zwischen Allianzpartner oder durch Vertragsbeziehungen erfolgen. Die Bildung eines Gemeinschaftsunternehmens, welches relativ zu den Allianzpartnern auf einer vor- oder nachgelagerten Wertkettenstufe anzusiedeln ist, bildet die dritte Möglichkeit zur Gründung einer vertikalen Allianz.

Abbildung 8.3 zeigt das Beispiel einer vertikalen Allianz mit wechselseitiger Kapitalbeteiligung

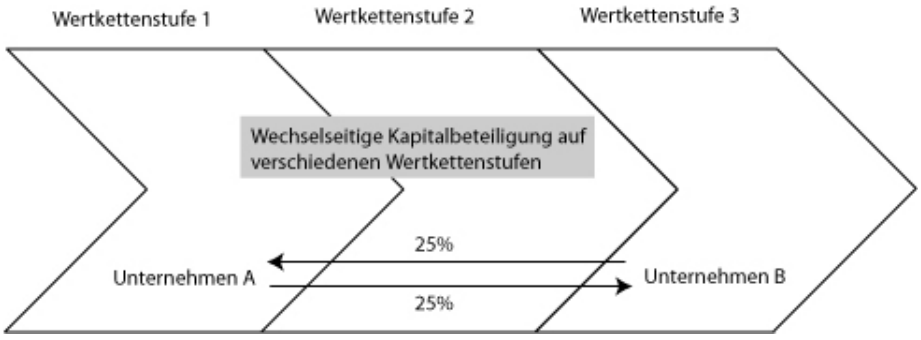

Abbildung 8.3: Wechelseitige Kapitalbeteiligung in vertikalen Allianzen

Diagonale Allianzen

Diagonale Allianzen sind Verbindungen zwischen Unternehmen, welche „weder auf dem gleichen relevanten Markt tätig sind (horizontal) noch in einem Käufer-Verkäufer-Verhältnis (vertikal) stehen.[9] Diagonale Allianzen können in Form eines Gemeinschaftsunternehmens, als wechselseitige Kapitalbeteiligungen oder als rein vertragliche Bindung gebildet werden. Als Beispiel seien Kooperationen unter Beteiligung von Netzbetreibern und branchenfremden Investoren, wie etwa Banken, zur Errichtung einer Übertragungsinfrastruk-

tur erwähnt. Obwohl beide Unternehmen in verschiedenen Branchen tätig sind, profitieren sie dennoch von einer Kooperation durch ein sog. Ressourcenpooling.

Juristischer Ansatz

Neben der Kategorisierung der verschiedenen Allianztypen im ökonomischen Ansatz muss das Verhältnis zwischen Allianzpartnern, das sog. Innenverhältnis, zwecks gegenseitiger Absicherung in juristischer Form festgehalten werden. Diese juristische Ausgestaltung hängt darüberhinaus von wirtschaftlichen und rechtlichen Rahmenbedingungen, wie z.B. steuerliche Erwägungen oder gesetzliche Regelungen, ab.

Horizontale Kooperationsverträge

Bei horizontalen Kooperationsverträgen handelt es sich um vertragliche Vereinbarungen ohne gesellschaftsrechtliche Verflechtungen.[10] Mögliche Formen solcher Kooperationsverträge sind: Abkommen über gemeinsame Forschungsvorhaben, Produktionsabkommen, Vertriebsvereinbarungen und Lizenzvergaben.

Vertikale Vertragsbeziehungen

Vertikale Vertragsbeziehungen stützen sich auf Verträge als Fundament für vertikale Kooperationen. Dabei gelten die Regeln einer vertikalen Allianz, nämlich die Verkäufer-Käufer Beziehung bzw. die Lieferanten-Abnehmer Verbindung. Dabei regeln die vertikalen Vertragsbeziehungen z.B. die Nutzung von Vertriebsstrukturen. Franchising, als Sonderform der Lizenz, stellt die intensivste vertikale Kooperationsform dar.

Gemeinschaftsunternehmen

Ein Gemeinschaftsunternehmen ist eine selbständige juristische Person, die durch mindestens zwei unabhängige Unternehmen gegründet wurde. Die Gründer beteiligen sich dabei kapitalmässig am neuen Gemeinschaftsunternehmen, was somit auch als Equity Joint Venture bezeichnet werden kann (equity = Eigen- bzw. Firmenkapital). Das Equity Joint Venture, wie bereits mehrmals in diesem Dokument angetroffen, stellt eine der intensivsten Formen strategischer Allianzen dar. Die Kapitalbindung ermöglich neben den „üblichen“ Vorteilen von Kooperationen auch eine intensive Zusammenarbeit im Management sowie einen Know-how Transfer. Dabei wird die Kontrolle von allen Allianzpartnern gemeinsam ausgeübt.

8.1.4 Wettbewerbspolitische Aspekte

In diesem Abschnitt werden die zuvor beschriebenen Eigenschaften von Allianzen in Bezug zur Marktstruktur analysiert. Dabei werden sowohl die wettbewerbsfördernden Vorteile als auch mögliche negative Effekte bezüglich Wettbewerbsverstössen angeschaut. Dabei spielt die ökonomische Erscheinungsform der strategischen Allianz eine wichtige Rolle und dient somit als Orientierung für die Gliederung der Unterkapitel in horizontalen und vertikalen Allianztypen.

Mögliche Auswirkungen horizontaler Allianzen

Wie oben bereits erwähnt sind horizontale Allianzen Verbindungen von Unternehmen, welche auf derselben Stufe der Wertschöpfungskette tätig sind. Juristische Erscheinungsformen beinhalten vertragliche Bindungen zwischen den horizontalen Allianzpartnern als auch Gemeinschaftsunternehmen. Hauptgründe für die Realisierung von horizontalen Allianzen sind economies of scale sowie eine Erhöhung des Marktanteils.

Mögliche positive Effekte

Strategische Allianzen können durch gemeinsame Finanzierung von Kapital Markteintrittsbarrieren überwinden (z.B. Lizenzgebühren). Somit wird eine Teilung der Amortisationskosten erreicht. Innovation wird nun durch horizontale Allianzen möglich.

Durch das Eintreten in einen neuen Markt wird die Konzentration desselben reduziert und die Wettbewerbsintensität steigt. Die Bildung horizontaler Allianzen fördert die obengenannte Zunahme von Teilnehmern in einem Markt und fördert somit den Konkurrenzkampf. Dabei besteht die Möglichkeit ein für die Allianzpartner neuer (geografischer) Markt zu erschliessen. Als Folge davon besitzten natürliche Monopole als günstigste Marktstruktur nur noch beschränkt Gültigkeit, da durch technologischen Fortschritt der Eintritt von neuen Marktbewerbern volkswirtschaftlich sinnvoll geworden ist. Die Duplikation von Investitionen und folglich die Verschwendung von Ressourcen fällt weg.

Ferner senken Economies of Scale Forschungs-, Entwicklungs-, Beschaffungs-, Produktions- und Marketingkosten. Diese positiven Auswirkungen können an Konsumenten weitergegeben werden. Es entsteht ein Preissenkungsdruck auf die übrigen Anbieter.

Schliesslich sorgt der Informationstransfer für den Austausch von Informationen und Know-how, was sich in einer Reduktion von Unsicherheit bemerkbar macht.

Mögliche negative Effekte

Die negative Seite von horizontalen strategischen Allianzen ist durch Marktmacht und Konzentration geprägt. Marktmacht kann definiert werden als: „...relativ grosser, von der Konkurrenz nicht kontrollierbarer Verhaltensspielraum eines Marktteilnehmers“ [11] Marktmacht führt zu überhöhten Preisen und folglich zu überdurchschnittlichen Gewinnen. Ein Allianzpartner, der bereits Marktmacht verfügt, kann durch eine horizontale Allianz diese Form von Macht noch vergrössern, was einen schlechten Einfluss auf die Kompetitivität ausübt. Dementsprechend können horizontale Allianzen die Konzentration auf einem relevanten Markt durch Gemeinschaftsunternehmen oder durch projektbezogene Zusammenarbeit auf derselben Wertkettenstufe erhöhen. Ein eventuelles Einschreiten der Wettbewerbsbehörden ist aber wahrscheinlich, was obengenannte Annahme als unrealistisch erscheinen lässt, da sie in Konflikt mit dem „Wesen“ einer Allianz, nämlich der Unabhängigkeit der Unternehmen ausserhalb des Allianzbereichs steht.

Bei der Kollusion stimmen die Anbieter ihre Aktionen so aufeinander ab, dass sie kollektive Vorteile erzielen (stammt aus der Oligopoltheorie). Mahrdt definiert Kollusion als „wettbewerbsbeschränkende Verhaltenskoordination, die implizit (tacit collusion) oder explizit (z.B. Kartelle) erfolgen kann“ ([4], S.162) Die Gefahr von Preiskollusionen (Vereinbarung

der Preishöhe) und Kapazitätskollusionen (Limitierung der Kapazität), z.B. beim Tätigen von erheblichen irreversiblen Inverstitionen mit anschliessendem Nachfragerückgang, kann daher zu Überkapazitäten führen.

Marktmächtige horizontale Allianzen können ferner in Märkten mit unreifem Preiswettbewerb negative Wirkungen hervorrufen.Z.B. können bei Forschungs- und Entwicklungsallianzen kommerzielle Innovationen zu Abmachungen führen, welche den Preis zu hoch ansetzen oder die Absatzmenge beschränken.

Als Beispiel für eine strategische Markteintrittsbarriere sei hier die Limitpreisstrategie (sog. Predatory Pricing) erwähnt. Dabei setzen horizontale Allianzpartner den Preis eines Produktes oder einer Dienstleistung so niedrig, dass potentielle Konkurrenten unmöglich kostendeckend in den Markt eintreten können (Gefahr eines möglichen Verstosses gegen die Wettbewerbsbestimmungen). Bei der Überkapazitätsstrategie hingegen kann eine eventuelle zusätzliche Nachfrage von den horizontalen Allianzpartnern zügiger und günstiger als von potentiellen Anbietern befriedigt werden.

Mögliche Auswirkungen vertikaler Allianzen

Vertikale Allianzen sind Verbindungen zwischen Unternehmen, die auf verschiedenen Stufen der Wertkette tätig sind und in einer Käufer-Verkäufer bzw. Zulieferer-Abnehmer Beziehung stehen. Vertikale Allianzen können sich in der Wertkette rückwärts (upstream) oder vorwärts (downstream) bewegen und „dienen der Reduzierung von Transaktionskosten, der Realisierung von technologischen Verbundvorteilen und marktstrategischen Zielsetzungen“([4], S.168)

Mögliche positive Effekte

Vertikale Allianzen ermöglichen einen effizienten Know-how Transfer, der die Unsicherheit über zu tätigende Inverstitionen unter den Allianzpartnern teilt. Als Absicherungsmassnahme kommt noch hinzu, dass sich vertikale Allianzen bei Änderungen der technologischen Rahmenbedingungen wieder auflösen lassen. Ferner besteht der Vorteil gegenüber dem Markt in der Geheimhaltungspflicht, welche durch vertragliche Verpflichtungen von unternehmenseigenem Know-how durch Allianzpartner geschützt wird. D.h. der Transfer von technologischem Know-how findet über vor- und nachgelagerte Wertkettenstufen, d.h. zwischen zwei relevanten Märkten statt, verlässt den „Allianzraum“ jedoch nicht.

Ein weiterer Vorteil von vertikalen Allianzen kann sich in Form einer effizienten Kundenorientierung bemerkbar machen. Synergieeffekte zwischen den Bereichen Koordination und Vermarktung ermöglichen eine effiziente Preisgestaltung was sowohl Privathaushalte als auch Firmenkunden profitieren lässt.

Mögliche negative Effekte

Vertikale Allianzen können durch die Errichtung von strukturellen bzw. strategischen Markteintrittsbarrieren den Zugang zu einem Markt soweit erschweren, dass es zu wettbewerbspolitischen Auseinandersetzungen kommen kann. Dabei können z.B. Inputfaktoren durch die Blockierung von ganzen Distributionskanälen verteuert werden. Dadurch

erhöht sich der Druck auf die Konkurrenz, ebenfalls mit Ressourcen aus beiden Stufen der Wertkette in den neuen Markt einzutreten.

8.2 Entwicklung der Telekommunikationsbranche

„Nach Ansicht der International Telecommunication Union (ITU) ist die Telekommunikationsbranche (...) die dynamischste der Erde.“ ([12], S. 5). Dass dieses Statement von 1997 stammt, mag zuerst erstaunen, könnte es doch genauso gut von heute stammen. Die Wachstumsprognosen von damals scheinen heute noch genauso zuzutreffen.

8.2.1 Ablauf der Argumentation

In diesem Teil der Arbeit soll ein Rückblick auf die Entwicklung der Telekommunikationsbranche gegeben werden, und es soll beleuchtet werden, wie die Entwicklung der Allianzen vonstatten ging. Aufgrund der Quellenlage kann angenommen werden, dass sich diese in 3 Phasen entwickelt haben.

1. Phase: Bereits zu Beginn der Liberalisierung des Telekommunikationsmarktes setzt eine Allianzbildung ein.

2. Phase: Die Allianzen entwickeln sich weiter, eine Aufgabenteilung setzt ein. Unternehmen, die sich ausschliesslich um den Mobilfunkbereich kümmern, entstehen in dieser Phase. Der Kampf um die UMTS Lizenzen, der zwischen der zweiten und dritten Phase stattfindet, heizt die Dynamik im Markt noch mehr an.

3. Phase: Heute treten vermehrt „multimediale“ Allianzen auf. Die Wertschöpfungskette hat sich verändert, die Telekommunikationsbranche ist auf Kooperationen der Unterhaltungsbranche und der Informatikbranche angewiesen (TIME Märkte). Da die Allianz- und Kooperationsforschung in diesem Bereich noch nicht sehr weit fortgeschritten ist, wird auch hier wieder auf Pressetexte und Onlinequellen zurückgegriffen. Im ersten Teil wird versucht, mit Hilfe der Literatur einen Einblick zu geben, wie die Liberalisierung des Marktes verlaufen ist, und welche Allianzen sich im Anschluss darauf ausgebildet haben. Die These geht davon aus, dass sich die staatlichen Player zuerst zusammenschlossen, um sich selber vor Konkurrenz zu schützen, aber auch, um den veränderten Kundenbedürfnissen Rechnung zu tragen, welche mit der Liberalisierung des Marktes aktuell geworden waren. Phase 2 ist zeitlich nicht genau abzugrenzen, anhand von Quellen wird versucht zu rekonstruieren, ob und wie sich die Allianzen weiterentwickelten und wie es zur Ausbildung des Mobilfunkbereich als eigenständiges Betätigungsfeld kam. Phase 3 hat schliesslich die UMTS- Technoloigie im Blickfeld. Ausgehend davon wird versucht, aufzuzeigen, wie es zu einer Rekonfiguration der Wertschöpfungskette kommt, und warum durch die zunehmende Konvergenz in Zukunft vermehrt mit multimedialen Allianzen zu rechnen sein wird.

8.2.2 Erste Phase: Beginn der Liberalisierung

Die Liberalisierung des Telekommunikationsmarktes ist wahrscheinlich der wesentliche Faktor, der die Unternehmen dazu zwingt, untereinander in Beziehung zu treten. Ebenfalls in diese Phase fällt der Eintritt branchenfremder Unternehmen in den Telekommunikationsmarkt. Die Marktdynamik während dieser Phase ist sehr hoch, was nun beleuchtet werden soll.

Von staatlichen Monopolisten zum freien Markt?

In diesem Abschnitt soll aufgezeigt werden, dass die rechtlichen, technischen und wirtschaftlichen Veränderungen ein Klima schufen, das die staatlichen Monopolisten schon zu Beginn der Liberalisierung dazu zwang, miteinander in Kooperation zu treten. Daher wird auf die rechtlichen, wirtschaftlichen und technischen Veränderungen im Markt näher eingegangen.

Die Entwicklung des Kommunikationsmarktes in Europa bis zur Liberalisierung

Lange Zeit wurde die Kommunikationsbranche als „natürliches Monopol“ betrachtet: Unter einem natürlichen Monopol versteht man Güterbereiche, in welchen der gesamtwirtschaftliche Ressourcenverbrauch dann am besten (und effizientesten) möglich ist, wenn ein einziges Unternehmen die Produktion übernimmt. (vgl. [12], S. 3) Dadurch fiel die Aufgabe, für die Kommunikation zu sorgen, dem Staat (als einzigem Anbieter) zu. Zudem bildete die Kommunikation mit anderen Verkehrsnetzen zusammen, die„infrastrukturelle Plattform“ ([12], S. 4) auf welcher „alle politischen, wirtschaftlichen und kulturellen Aktivitäten eines Landes stattfanden“Somit übernahm der Staat die Rolle des Installateurs und Betreibers von Kabelnetzen, was zu lukrativen Einnahmen führte. Weiter operierte der Staat als „Transporteur“ und war mit eigenen Telefon- und Telefgrafiegesellschaften am Markt tätig. Diese Gründe führten zur Bildung staatlicher Monopole in den Bereichen Telegraphie und Telephonie.

Diese Situation änderte sich, als die EG Kommission in ihrem Grünbuch von 1987 eine Liberalisierungsdebatte auslöste ([13], S. 25). Diese führte zur schrittweisen, 1996 abgeschlossenen, Liberalisierung des Marktes. Die rechtlichen Schwierigkeiten bestanden bei diesem Vorgehen darin, zu garantieren, dass rechtliche „Leitplanken“ geschaffen wurden, um Missbrauch der bisherigen Telekommunikationsanbieter auszuschliessen. Dies war aus verschiedenen Gründen nicht einfach:„Die grösste Gefahr für die Entstehung und Sicherung tatsächlichen Wettbewerbs geht dabei von den ehemals rechtlich geschützten Monopolisten in den europäischen Mitgliedsstaaten(...) aus. Da der erstrebte Wettbewerb zu Lasten ihrer Marktanteile geht, haben sie den grössten Anreiz diesen zu blockieren. Darüber hinaus verfügen sie über die grössten Potentiale, um sich des Wettbewerbs auch mit unfairen Mitteln zu erwerben.“ ([13], S. 26).Weiter weist Jaggi darauf hin, dass diese Akteure mit ihren ausgebauten Kabelnetzen bereits über eine Infrastruktur verfügen, welche die neu eintretenden Player benützen mussten, um überhaupt Dienste an die Endkunden bringen zu können: Die Kontrolle dieser so genannten„Bottleneck - Einrichtungen“ ermöglicht es den staatlichen Anbietern, weiterhin mit Monopolgewinnen rechnen zu können.

Somit bestand eine Gefahr, dass die staatlichen Akteure den liberalisierten Wettbewerb ausschalten könnten. ([13], S. 26).

Entwicklung des Mobilfunks bis zum GSM- Netz

Die technische Entwicklung der mobilen Kommunikation nahm jedoch schon vor der Liberalisierungsdebatte ihren Anfang: Vor allem der Mobilfunkbereich machte eine rasante Entwicklung durch, die schon in den fünfziger Jahren einsetzte: So fügte Deutschland im Jahr 1958 die bisher existierenden Funknetze zu einem einheitlichen Funknetz zusammen („Öffentliches bewegtes Landfunknetz (öbl)A“ zusammen. Gegen Ende der fünfziger Jahre war dieses A - Netz das bedeutendste öffentliche Mobilfunknetz der Welt (die Flächenabdeckung betrug um die 80%). Dennoch fiel die Nutzung des Netzes gering aus, was vor allem auf die hohen Kosten zurückzuführen ist. (Grundgebühr: 270 Mark pro Monat).

1972 startete dann das B-Netz. Eine neue technische Errungenschaft bestand darin,dass die Teilnehmer im B-Netz selbst wählen konnten, und nicht mehr auf die Vermittlungangewiesen waren. Voraussetzung war jedoch immer noch, dass man die Zone kannte, in welcher sich der Gesprächspartner aufhielt (Deutschland war in 150 solcher Zonen aufgeteilt).Auch waren Telefonate ins Ausland über dieses Netz möglich.

Das C Netz, welches 1985 in Betrieb genommen wurde, übertrug die Sprachsignale zwar noch analog, die Steuersignale jedoch digital ([14]).

Die Entscheidung, einen weltweiten Standard für die Mobilkommunikation durchzusetzen, fand bereits im Juni 1982 statt. Im Vergleich zu den vorherigen Standards war der Global System for Mobile Communication (GSM) digital, und enthielt eine grössere Bandbreite. Am 7. September 1987 schlossen sich Netzbetreiber, welche die Wichtigkeit eines Standards erkannten, zur GSM Mol Gruppe zusammen. 1991 wurden die ersten GSM Netze aufgeschaltet, die nun länderübergreifend miteinander kompatibel waren. Auch konnten erstmals Kurznachrichten über das GSM Netz versendet werden. ([15]) Ebenfalls durch die technische und wirtschaftliche Entwicklung angestossen, veränderten sich die Kundenbedürfnisse: Ein weiteres Kriterium, welches die Allianzenbildung vorantrieb, waren sog. „one-stop-shoprequirements“ welche durch die zunehmende Globalisierung verursacht wurden. Dabei verlangen Unternehmen auf ihre Bedürfnisse zugeschnittene Dienste, um ihre (zunehmend globalen) Aktivitäten zu koordinieren. Diese Unternehmen verfügen über grosse Nachfragemacht. ([12], S.5).

Die Teilnehmer der ersten Phase

Sowohl die staatlichen, als auch die nichtstaatlichen Player haben in der ersten Phase mit jeweils unterschielichen Problemen zu kämpfen:

Probleme der staatlichen Unternehmen

Zwar besassen die staatlichen Player, welche über die Bottleneck Einrichtungen verfügten, einige Vorteile, doch hatten auch sie mit spezifischen Problemen zu kämpfen; da sie vor der Aufgabe standen, einen umfassenden Transformationsprozess einzuleiten. Themen wie Kostensenkungen oder Restrukturierungen traten nun in den Vordergrund, welche vorher

weniger wichtig waren. Ebenfalls spürbar dürfte der einsetzende Marktdruck gewesen sein, konnten doch Kunden nun damit drohen, den Anbieter zu wechseln. Daher waren die ehemaligen Monopolisten dazu gezwungen, ihre Kommunikationsanlagen zu modernisieren, um mit neu eintretenden Playern Schritt halten zu können.

Strategien der neu in den Markt eintretenden Unternehmen

Bei den neu eintretenden Playern lassen sich verschiedene Strategien ausmachen, um in den bestehenden Telekommunikationsmarkt einzudringen. Diese können auch kombiniert zur Anwendung gelangen. Soll ein flächendeckendes Universalangebot errichtet werden, wird zunächst versucht, ein landesweites Netz aufzubauen. Der Fokus liegt dabei zunächst auf Geschäftskunden, später wird versucht, eine Ausweitung auf alle Kundensegmente vorzunehmen. Als alternative Strategie bietet sich an, lediglich die Infrastruktur anzubieten, und auf Dienste für Endkunden zu verzichen (Carriers Carrier). Ein möglicher Ansatz ist auch, zunächst nur Grossstadtregionen abzudecken, und sich später auf dieVerbindungen zwischen diesen zu konzentrieren. Dabei werden die Bedürfnisse einer engdefinierten Kundengruppe (wie z.B. der Finanzwelt) in den Vordergrund gerückt, und erst später wird versucht, auch andere Kundensegmente abzudecken. Beim Privatkundenfokus lassen sich bereits bestehende Netze, wie z.B. Kabelfernsehnetze, weiter verwenden, um sich Zugang zu den Kunden zu verschaffen.

Probleme der neu in den Markt eintretenden Unternehmen

Den neu eintretenden Playern fehlt zumeist das technische, als auch das praktische marktbezogene Wissen, das sie erst langsam erwerben müssen. Als hemmend kann sich auch auswirken, dass die neuen Player mit teilweise raschem Wachstum fertig werden müssen, und darum mit der Notwendigkeit konfrontiert werden, passende Ablaufsstrukturen zu schaffen. Bei den Hauptproblemen für die neuen Player nennt Lechner ([12], S. 11) die Bottleneck-Problematik: „Zuletzt treten immer wieder ungelöste regulatorische Fragen auf, die ihre Aktivitäten und die Rentabilität des Geschäftes ernsthaft behindern. Zu nennen ist hier beispielsweise die Frage, welche Nutzungsgebühr sie für den Zugang zu den Netzen des ehemaligen Monopolisten zu entrichten haben."

Wichtige Allianzen der ersten Phase

So ist es auffällig, wie viele Allianzen sich gerade zu Beginn der Marktliberalisierung zwischen den staatlichen Playern bildeten: Die strategische Allianz„Atlas" zwischen Deutsche Telekom und France Telekom im Jahr 1996, oder „Unisource" an welcher ebenfalls die Schweizerischen PTT Betriebe zusammen mit der niederländischen PTT Telecom BV, und Telia AB (der schwedischen Telekommunikationsgesellschaft). Uniworld wiederum war das Ergebnis einer Verbindung der bestehenden Allianz Unisource mit AT&T, welche 1997 stattfand. Die Allianzen der ersten Phase haben dann in der zweiten Phase mit dem sich verändernden Markt zu kämpfen, und lösen die ursprünglichen Allianzen auf, oder gehen zusätzlich neue Allianzen ein. Dies wird am Beispiel Unisource in Phase 2 genauer beleuchtet.

Zusammenfassung der ersten Phase

Die Allianzenbildung in der ersten Phase wurde vor allem von den (staatlichen und nichtstaatlichen) Unternehmen vorangetrieben, um sich den Herausforderungen der Liberalisierung des Marktes in Europa zu stellen. Staatliche Player waren gezwungen, sich mit andern staatlichen Playern zusammenzuschliessen, um den geänderten Kundenbedürfnissen Rechnung zu tragen. Bereits zu Beginn der Liberalisierung des Kommunikationssektors setzte also eine rege Allianzbildung ein. Die Allianzen sind nicht nur ein spezifisches Phänomen der modernen Mobilfunkbranche, sondern sind im gesamten Kommunikationsbereich zu finden. Die Bildung von Allianzen ist also kein Phänomen allein der Mobilfunkindustrie, wie es heute im Rückblick erscheinen könnte, sondern durchzieht während der Marktöffnung die gesamte Telekommunikationsbranche.

8.2.3 Zweite Phase: Wachstum und Differenzierung

Die zweite Phase lässt sich von der ersten Phase nicht durch einen klaren Bruch abgrenzen, vielmehr kommt es zu einer Veränderung der Allianzen. Diese ist durch das rasche Wachstum der Allianzpartner bedingt. An einem Beispiel soll diese Entwicklung im zweiten Teil des Abschnitts verdeutlicht werden. Wie diese verlief, ist Thema dieses Abschnittes.

Der Übergang von der ersten zur zweiten Phase

Die erste Phase ist geprägt gewesen vom Zusammenschluss mehrheitlich staatlicher Player. Der Mobilfunkmarkt an sich hat eher eine untergeordnete Rolle gespielt, obwohl er zu diesem Zeitpunkt als durchaus zukunftsträchtig angesehen wird ([12], S.5). Erst in der zweiten Phase kommt es dann innerhalb der Allianzen zu einer Ausdifferenzierung der einzelnen Geschäftsfelder. Dies hängt mit der zunehmend komplexer werdenden technischen Entwicklung zusammen, als auch mit dem rapiden Wachstum der bestehenden Allianzen, welches zu klaren Organisationsstrukturen zwingt. Dadurch kristallisiert sich nun der Mobilfunkmarkt als eigenständiger Geschäftsbereich heraus und die Allianzen bilden spezialisierte „Einheiten“ um diesen Bereich zu bearbeiten. Indem die Allianzen wachsen, ist es ihnen auch zunehmend möglich, die gesamte Wertschöpfungskette zu kontrollieren. Der Übergang von der ersten zur zweiten Phase lässt sich nicht fest abgrenzen, sondern gestaltet sich fliessend.

Beispiel Unisource/Uniworld

Unisource ist ein Joint Venture, das zu Beginn der 90er Jahre aus der schwedischen Telefongesellschaft Telia und der niederländischen Telefongesellschaft KPN hervorgegangen ist. Schweden ist zu Beginn der 90er Jahre eines der wenigen Länder gewesen, das über kein formales Monopol in der Telekommunikation verfügt hat. Prinzipiell ist der Markt somit für Mitbewerber geöffnet. Dennoch nimmt die Telefongesellschaft Telia faktisch eine

Monopolstellung ein, da kein anderes ausländisches Unternehmen auf dem Markt operiert , ein Indiz für die damals vorherrschende Branchenlogik, sich ausschliesslich auf den eigenen Heimmarkt zu konzentrieren.

Um 1990 ändert sich diese Situation: Mit der sich abzeichnenden Marktliberalisierung sieht Telia einen starken Wettbewerb auf sich zukommen, da Schweden keine Gesetze kennt, um dies zu verhindern: „Wettbewerber sind zwar noch nicht im Markt, stehen jedoch bereits an dessen ‚Pforten' die nicht durch staatliche Auflagen geschützt sind." ([12], S.123). Selber kann sie aber nicht auf andern Märkten aktiv werden, da diese staatlich geschützt sind. (vgl. [12], S. 123). Daher werden Gespräche mit der niederländischen staatlichen Gesellschaft KPN aufgenommen.

1992 wird ein Joint Venture namens „Unisource" gegründet, das aus zwei operativen Einheiten, welche in Schweden und den Niederlanden angesiedelt sind, besteht ([12], S. 134). In der ersten Phase bleibt das Joint Venture Unisource jedoch erstmal eine „ juristische Hülle" ([12], S. 140), und dient dem Zweck, dass sich die beiden Firmenkulturen kennenlernen. Ebenfalls 1992 tritt die Schweizer Telekom (die mit der Marktöffnung 1997 in Swisscom umbenannt wird) der Allianz bei. Ziel der Schweizer Telekom ist es, international tätig werden zu können.

1994 tritt ebenfalls die spanische, staatliche „Telefonica" der Unisource Allianz bei. Zwischen 1993 bis 1995 kommt es zur Ausbildung mehrerer operativer Einheiten, wobei auch die Ziele der Allianz, welche vorher eher vage formuliert gewesen sind, klarer werden: „Mehrere Einheiten werden neu gegründet, wodurch eine Unternehmensgruppe entsteht, die mit sechs Geschäftseinheiten auf insgesamt drei Geschäftsfeldern tätig ist." ([12], S. 154). 1994 kommt dann ein weiteres Geschäftsfeld dazu, die „Personal Services" welches selbst wiederum aus zwei Einheiten besteht: „Card Services (UCAS)" welche sich um die Entwicklung von Telefonkarten kümmert, und „Mobile Services (UMS)" welche mobile Mehrwertdienste in Europa offeriert. Gleicheitig entwickelt sich ebenfalls ein enger Kontakt zwischen Uniworld und der amerikanischen Firma AT&T. AT&T verfolgt die Absicht, ihre europäischen Aktivitäten im Rahmen einer gesamteuropäischen Lösung zu verbinden. Es folgt eine Zusammenarbeit, um einheitliche europäische Datendienste für Unternehmen anbieten zu können. Daraus entwickelt sich später wiederum ein Joint Venture zwischen Unisource und AT&T mit dem Namen „Uniworld." Unisource beteiligt sich an der „World Partner Organisation" welche ein lockerer Verbund von Telekommunikationsunternehmen darstellt, der seit 1993 besteht, um weltweit ein einheitliches Datennetz garantieren zu können. ([12], S. 173f). 1995 wird erneut eine Neustrukturierung vorgenommen: Unisource und AT&T arbeiten einen Plan aus, nach welchem drei neue Joint Ventures gegründet werden.

- „AT&T Unisource Communication Services„ (AUS) kümmert sich darum, pan- europäische Kommunikationsdienste anzubieten.
- „AT&T Unisource Multimedia & Solutions" (AUM) kümmert sich um Internet-Consulting und Outsourcing-Dienste.
- „AT&T Unisource Participation Company" (AUP) organisiert den Erwerb von Minder- und Mehrheitsbeteiligungen in andern Ländern, kümmert sich also darum, geeignete

> Distributoren in Ländern, in welchen die Allianz noch nicht vertreten ist, zu finden. ([12], S. 185)

Damit wird bezweckt, die Wertschöpfung zu optimieren: „Mit der Neustrukturierung wird eine Aufteilung der Wertschöpfungskette entlang von drei Ebenen vorgenommen. Auf einer ‚Bit-Schaufel' die als Carrier's Carrier über terrestrische, mobile und satellitengestützte Netze verfügt und für den Transport digitaler und analoger Nachrichten Informationen zuständig ist, setzen auf einer zweiten Ebene Einheiten auf, die Sprach- und Datendienste anbieten (...). Multimediale Dienste bilden zuletzt die dritte Ebene der Wertschöpfungskette (...)." ([12], S. 185).

Zusammenfassend lässt sich sagen, dass während der zweiten Phase die Allianzen mit rapidem Wachstum zu kämpfen haben. Zudem erfordert es die zunehmende technische Komplexität, dass innerhalb der Allianz eine Arbeitsteilung einsetzt, und Spezialgebiete erforscht werden. Damit werden auch die auch die Ziele der Allianz klarer umrissen. Die Beziehungen inner- und ausserhalb des Allianzsystems werden jedoch komplizierter und schwieriger zu durchschauen, da die Joint- Ventures ihrerseits selbst wieder Allianzen eingehen.

8.2.4 Dritte Phase: multimediale Allianzen

Die zweite Phase hat sich durch eine zunehmende Aufgabenteilung innerhalb der Telekommunikationsunternehmen ausgezeichnet, welche nun weitestgehend abgeschlossen ist: In der dritten Phase wetteifern verschiedene Mobilfunkunternehmen miteinander, die einerseits noch immer in Allianzen mit ihren ursprünglichen Telekommunikationsfirmen stehen, andererseits auch selber untereinander Allianzen eingehen. Ziel der Allianzen in der dritten Phase ist es, weitergehende Dienste anbieten zu können, die man selber nicht anbieten kann, und den Kunden stärker an das Unternehmen zu binden. Wesentlich zu dieser Entwicklung beigetragen hat die UMTS- Technologie.

UMTS

UMTS ist die englische Abkürzung für „Universal Mobile Telecommunications System" und bezeichnet den europäischen Standard der dritten Mobilfunkgeneration, der international den Namen IMT-2000 trägt. UMTS garantiert eine höhere Sprachqualität, als bei den Mobilfunktelefonen der zweiten Generation. Zusätzlich, und das ist der wichtigere Grund, verspricht UMTS „völlig neue Anwendungen mit hoher Dienstgüte" zu eröffnen. Dies wird durch die höhere Datenübertragungsrate erreicht (5000 kHz pro Kanal, anstelle von 200 kHz wie bei GSM). Mobile Internetzugänge und damit verbundener E-Commerce bieten damit nur einen kleinen Ausschnitt der Möglichkeiten: Zusätzlich lassen sich auch „Entertainment Inhalte, wie vollständige Musikstücke, Videoclips und netzwerkfähige Spiele (...) mit Hilfe der UMTS- Technologie auf mobilen Endgeräten realisieren." ([16]).

Das bedeutet, es kommt zu einer Ausweitung der Wertschöpfungskette. Beschränkte sich der Mobilfunkmarkt vorher auf den Infrastrukturanbieter, den Netzanbieter und den Endgerätemarkt, so gibt es nun zusätzlich Platz für Content- Anbieter und Dienstanbieter. In der Hoffnung, solcherart einen Mehrwert generieren zu können, tätigen die Kommunikationsunternehmen grosse Investitionen. Um das Jahr 2000 werden in den meisten europäischen Ländern UMTS- Lizenzen von staatlichen Institutionen vergeben, welche sicherstellen sollen, dass die Unternehmen, welche eine Lizenz erwerben, auch in der Lage sind, die geforderte Netzabdeckung zu erreichen. In der Schweiz beispielsweise übernimmt die Regulierungsbehörde ComCom diese Aufgabe (vgl. [17]). Den Zuschlag erhalten die meistbietenden Unternehmen. Bereits im Vorfeld der Lizenzierung kommt es zu Erschütterungen der bestehenden Allianzen, und es werden neue Allianzpartner gesucht, um die hohen Kosten tragen zu können. Besonders in Deutschland nehmen die gebotenen Summen horrende Ausmasse an. Dies hat Auswirkungen auf die zu diesem Zeitpunkt existierenden Allianzen, und führt zu neuen Kooperationen, wie der Neuen Zürcher Zeitumg vom 19.08.2000 zu entnehmen ist: „Auch der Hongkonger Mischkonzern Hutchison Whampoa war mit den hohen Kosten (...) sehr unzufrieden. Er sprengte noch am Donnerstagabend die erst vor wenigen Wochen mit der niederländischen KPN geschlossene Allianz (...). KPN, die Anfang dieses Jahres die Mehrheit am drittgrössten deutschen Mobilfunkanbieter E-Plus von Veba und RWE erworben hat, will die Lizenzgebühren nun gemeinsam mit dem amerikanischen Partner Bell South und der zu 15% an der KPN-Mobile-Tochter beteiligten NTT Docomo aus Japan tragen. Hutchison denkt unterdessen, wie mitgeteilt wurde, über alternative Wege auf den deutschen Mobilfunkmarkt nach, wie das Leasing von Frequenzen oder das Anbieten von Dienstleistungen.“ ([18]).

Da die Lizenz- und Investitionskosten für den Netzausbau enorm hoch sind, steht die 3. Mobilfunkgeneration heute unter Druck, die Einnahmen pro User zu erhöhen. Deshalb wird intensiv nach neuen Diensten gesucht, um die Ausgaben zu amortisieren. ([19]). Dadurch muss mit branchenfremden Unternehmen zusammengearbeitet werden. Es entstehen neue Märkte.

TIME - Märkte

Unter TIME - Märkten wird das Zusammenwachsen von Unternehmen der Telekommunikations- und Informatikbranche sowie der Medien- und Entertainment- Industrie verstanden. ([20], S.1). Voraussetzung für dieses Zusammenspiel der unterschiedlichen Branchen ist die technische Konvergenz: Diese stellt die Basis für alle anderen Konvergenzprozesse dar. Vorangetrieben wird technische Konvergenz durch die Erhöhung der Bandbreite und Prozessorenleistung, zum andern durch die Etablierung plattformübergreifender Standards und Protokolle (vgl. [20], S. 2). Die ökonomische Konvergenz basiert auf der Rekonfiguration von Wertschöpfungsketten. So werden die Grenzen bislang klar getrennter Bereiche, wie z. B. zwischen Telekommunikations- und Medienunternehmen unscharf, da Nachrichten auf unterschiedliche Endgeräte versandt werden können; Telekommunikationsanbieter könnten zu Nachrichtenanbietern werden. ([20], S. 2). Weiter entstehen vor- und nachgelagerte Branchen, und es entstehen neue Wettbewerbsbeziehungen. Der Erfolg dieser Angebote hängt letztlich aber immer noch von der Konvergenz der Gebrauchsweisen ab, das bedeutet, die Endkunden müssen die Angebote auch nutzen. Denn erst der Wille des Nutzers,

diese Angebote in seinen Medienkonsum aufzunehmen, ist die notwendige Bedingung für den Markterfolg. ([20], S. 3). Hier verhalten sich die Nutzer zur Zeit noch widerstrebend, was mit der zunehmenden technischen Komplexität der Angebote zu tun haben könnte: „Es zeigt sich jedoch, dass die Mediennutzung sich sehr viel langsamer und geringfügiger verändert, dass die Rezipienten sich in ihrer Medienwahl sehr viel konservativer verhalten, als von den Akteuren der TIME- Industrie gewünscht.“ ([20], S. 3).

Beispiele von multimedialen Allianzen finden sich zur Zeit häufig, sind aber keineswegs auf den Mobilfunkmarkt beschränkt, so gab AOL beispielsweise bekannt, mit Philipps zusammenzuarbeiten, um iTV- Settopboxen herzustellen, um TV- Angebote übers Internet zu beziehen. ([21]) Motorola dagegen plant, iTunes von Apple auf das Handy zu übertragen, damit über UMTS Musikdateien auf das Handy übertragen werden können. ([22]) Eine Allianz mit einer Firma, welche ehemals aus dem Printbereich stammt, findet sich bei Swisscom: Publicitas, die auch heute noch im Inserategeschäft tätig ist, bietet nun nun ihrerseits „Wettbewerbe, Info- und Chat-Services, Abstimmungen und demnächst MMS-basierte Services nach Themen“ an. ([23]) Ein Beispiel für eine Allianz, welche die technische Konvergenz voranzutreiben sucht, ohne die, wie oben dargestellt, weiterführende Services kaum möglich sind, ist die im nächsten Abschnitt vorgestellte Open- Mobile Alliance, welche sich um die Entwicklung und Durchsetzung von Standards kümmert.

8.3 Aktuelle Allianzen in der Mobilfunkbranche

8.3.1 Mobilfunk-Allianzen

Wie vorher erklärt wurde, gibt es unterschiedliche Arten von Allianzen, sprich vertikale, horizontale und diagonale. Auch im Mobilfunk gibt es verschiedenste Allianzen. Die Art der Allianzen, welche aus Endkundensicht am stärksten wahrnehmbar sind, sind die horizontalen Allianzen zwischen mehreren Mobilfunkanbietern aus verschiedenen Ländern. Zwei aktuelle Beispiele sind die Starmap Mobile Alliance sowie die FreeMove Alliance, auf diese wird im Folgenden noch eingegangen. Naheliegend sind sie, weil ihre Entstehung und Entwicklung meist unbewusst wahrgenommen wird, z.B. durch Werbung oder Ankündigungen der Mobilfunkbetreiber selbst. Alle Besitzer von mobilen Endgeräten haben schon vom Begriff Roaming gehört, oder sogar schon einmal davon Gebrauch gemacht. Roaming steht bei den meisten vertikalen Allianzen momentan im Vordergrund. Natürlich arbeiten sie auch noch an anderen Dienstleistungen für Geschäfts- und Privatkunden. Generell ist es schwierig über den bisherigen Erfolg dieser Allianzen zu sprechen, da viele erst vor kurzem gegründet wurden. Das bedeutet, dass viele Allianzen bis jetzt kaum neue Dienstleistungen anbieten, und falls doch, sind es bei allen Allianzen ähnliche Dienste.

Vertikale Allianzen im Mobilfunk sind weniger bekannt, weil Endbenutzer nicht direkt von den Kooperationen profitieren, sondern eher indirekt. Vertikale Allianzen bestehen vorwiegend zwischen Chipherstellern und Geräteherstellern oder Mobilfunkanbietern und Contentprovidern. In dieser Arbeit werden keine Beispiele von vertikalen Allianzen näher geschildert, da sie zu wenig Relevanz zu den vorherigen Ausführungen haben.

Diagonale Allianzen können definitionsgemäss nicht als Beispiel betrachtet werden, da es sich beim Mobilfunk um eine einzige Branche handelt. Unternehmen bzw. Institutionen orthogonal verlaufender Branchen bilden diagonale Allianzen.

Als eine besondere Form von Allianzen gilt die Kombination von vertikalen und horizontalen Allianzen, wie zum Beispiel die Open Mobile Alliance. Sie setzt sich aus verschiedenen Unternehmen aus allen Teilen der Wertschöpfungskette zusammen. Eine Eigenschaft dieser Allianzen sticht heraus: die Zusammenarbeit konkurrenzierender Unternehmen. Das ist auf den ersten Blick nicht ganz verständlich, hat aber einige Vorteile für die Beteiligten sowie auch für die Endbenutzer. Die Open Mobile Alliance wird später genauer betrachtet.

Bei der Gründung solcher Allianzen, wie im ersten Teil erwähnt, stehen immer mehr oder weniger die gleichen Ziele im Vordergrund. Das gilt auch im Mobilfunk. Hier sind die Motivationen, um vertikale Allianzen zu gründen vor allem durch die Vereinheitlichung der Benutzung zu erklären. Alles soll im Netzwerk von anderen Mobilfunkanbietern, generell solche aus anderen Ländern, genau so funktionieren, wie im Heimnetz. Eine spezielle Motivation ergibt sich aus der Marktmacht von Vodafone. Vodafone, als einer der ersten und heutzutage grössten Mobilfunkanbieter Europas oder sogar weltweit, ist für viele kleinere Anbieter zu stark im Markt vertreten. Deshalb haben sie sich zu Allianzen zusammengeschlossen, um Vodafone ein ebenbürtiger Konkurrent zu sein und ihn sogar zu übertreffen.

8.3.2 Allianzen auf Provider-Ebene

Vodafone-Gruppe

Vodafone gehörte im Jahre 1992 zu den weltweit ersten Netzbetreibern, welche die GSM-Technologie einsetzten und ihr zum Durchbruch verhalfen. Heute betreiben sie eines der grössten und leistungsstärksten GSM-Mobilfunknetze der Welt. Das Unternehmen ist in 26 Ländern über 5 Kontinente vertreten. In etwas mehr als 50% der Fälle sind sie Eigentümer der jeweiligen Mobilfunkanbieter, in den anderen Fällen haben sie Anteile an den sonst eigenständigen Anbietern. Hinzu kommen noch Partnernetzwerke in zusätzlichen 14 Ländern. Vodafone hat nur schon in den primären 26 Ländern ca. 146,7 Millionen Kunden.

Vodafone hat in jüngster Zeit wieder eine Vorreiterrolle eingenommen, und zwar in Bezug auf UMTS. Sie haben eine „Mobile Connect Card“ entwickelt, mit welcher es möglich ist, mit einem Laptop über das UMTS Netzwerk ins Internet zu gelangen. Sie wurde primär den Geschäftskunden angeboten. Auch das Portal VodafoneLive wurde im Zuge des UMTS Netzes entwickelt. Zu Beginn war das Portal nur mit GSM-Geräten zugänglich, nun wird parallel der Zugang für UMTS-Geräte ermöglicht. Vodafone hat schon einige UMTS fähige Geräte auf den Markt gebracht.

In der Schweiz hat Vodafone einen 25 prozentigen Anteil an Swisscom Mobile. Das lässt sich unschwer erkennen, da die Swisscom Mobile schon seit einiger Zeit Geräte vertreibt, welche den Dienst VodafoneLive anbieten.

Abbildung 8.4 zeigt die weltweite Abdeckung von Vodafone. Um die Weltkarte herum ist der Anbieter des jeweiligen Landes aufgelistet. [24]

Abbildung 8.4: Weltweite Abdeckung von Vodafone

StarMap Mobile Alliance

Die Starmap Mobile Alliance wurde im Oktober 2003 von 9 grösseren Mobilfunkanbietern Europas gegründet. Dazu gehören: Amena (Spanien), O2 (Deutschland, UK und Irland), One (Österreich), Pannon GSM (Ungarn), sunrise (Schweiz), Telenor Mobil (Norwegen) und Wind (Italien). Sie hielten damit insgesamt mehr als 40 Millionen Kunden. Im März 2004 wurde die Allianz mit dem dänischen Anbieter SONOFON SA ergänzt. Und im September 2004 wurde der elfte und vorläufig letzte Mobilfunkanbieter, Eurotel (Tschechische Republik) in die Allianz aufgenommen. Mittlerweile hat die Starmap Mobile Alliance insgesamt 53 Millionen Kunden. Das ist zwar immer noch nur rund ein Drittel der Kunden, welche von Vodafone erreicht werden.

Geschäfts- sowie Privatkunden können bereits von verschiedenen Diensten profitieren. Zu den verfügbaren Diensten gehören unter anderem GPRS und MMS Roaming. Die Kunden haben die Möglichkeit von jedem Gebiet des jeweiligen Allianzmitgliedes GPRS und MMS Dienste zu nutzen, als würde auf die Dienste vom Heimnetz zugegriffen werden. Zudem sind Voice-Mail und Kurznummern länderübergreifend gültig, was Vorteile für Reisende bietet. Alle Mobilfunkbetreiber innerhalb der Allianz streben einen gemeinsamen Nenner in Bezug auf Quality of Service (QoS) an.

Im Dezember 2004, also erst vor kurzem, wurde der „Multi Country Corporate Service“ gestartet. Dieser Dienst hat zum Ziel, dass generell vereinfachte Verträge vorgelegt werden und einheitliche Geschäftsbedingungen in allen Allianznetzen gelten. Vorläufig ist diese Dienstleistung nur den Geschäftskunden vorbehalten. Unternehmen, welche sich für den „Corporate Service“ entscheiden, erhalten einerseits die Möglichkeit Informationen betreffend den Ausgaben für Telefonate und Daten jederzeit am Computer abzufragen. Andererseits wird ein neues länderübergreifendes Rabattsystem eingeführt, mit welchem die Kunden zusätzliche Vergünstigungen auf die landesüblichen Tarife erhalten. Die genaue Funktionsweise wurde aus den Unterlagen nicht ersichtlich.

Zusätzlich zu den angebotenen und geplanten Diensten besteht eine Kooperation der Allianzmitglieder mit dem Ziel, 3G (UMTS) Handhelds zu planen und zu entwickeln. Ein erster Handheld wurde schon zum Vertrieb freigegeben, der Xda II Pocket PC. Das Gerät soll Kunden den Zugang zu Videos, Bilder und Applikationen ermöglichen. Wobei zu sagen ist, dass sunrise dieses Gerät nicht im Angebot hat. [25]

FreeMove (Alliance)

Im Juli 2003 hat T-Mobile gemeinsam mit Orange SA, Telefónica Móviles, und Telecom Italia Mobile (TIM) eine Mobilfunkallianz gegründet. Unter dem Namen FreeMove haben sie in 21 Ländern Europas insgesamt fast 170 Millionen Kunden, was sie zum grössten Mobilfunkanbieter macht, auch grösser als Vodafone. Weltweit zählt die FreeMove Allianz rund 230 Millionen Kunden. [26]

Das übergeordnete Ziel von FreeMove ist, dass das mobile Endgerät von überall genauso verwendet werden kann wie man das von zu Hause gewohnt ist. Deshalb auch der Satz: „Discover mobility without complexity“ oder die Bezeichnung „Virtual Home Environment.“ Weitere Ziele und Zukunftsvisionen von FreeMove betreffen vor allem die höhere Wettbewerbsfähigkeit im Markt, integrierte und einheitliche Dienste für Kunden und massive Kosteneinsparungen durch die gemeinsame Beschaffung von Endgeräten.

Sie planen eine Art „one-stop-shop“ anzubieten, bei welchem den Kunden ein zentrales Account Management zur Verfügung steht. Wie die anderen Allianzen stellt auch die FreeMove Allianz die Roaming Dienste ins Zentrum ihrer Bemühungen. Betreffend Roaming sollen die Preise transparent sein und es soll ein intuitives System von geografischen Zonen entstehen. Welches Netzwerk oder zu welcher Zeit die Dienste eines Mobilfunkanbieters genutzt werden, soll keinen Einfluss haben.

FreeMove unterteilt ihre Dienste in zwei grobe Kategorien, solche für Privatkunden und solche für Geschäftskunden. Roaming ist in beiden Kategorien vertreten. Nicht nur die Gewährleistung, jederzeit unter derselben Nummer erreichbar zu sein, sondern auch das Anzeigen der Nummer des Anrufenden, wie das im Heimnetz der Fall ist. Neben den eher sprachorientierten Diensten für Privatkunden, stehen bei den Geschäftskunden ebenso datenorientierte Dienste im Mittelpunkt. Zum Beispiel Roaming für GPRS Übertragung bzw. Nutzung. Des Weiteren soll eine international zentralisierte Instanz für Reporting und Analyse eingerichtet werden, um Geschäftskunden eine transparente Sicht auf Nutzung und Kosten zu garantieren.

Die FreeMove Allianz beteiligt sich mit ihren Diensten noch an grossen Events und Veranstaltungen . Zum Beispiel organisierte sie bei den olympischen Spielen in Athen spezielle Mobilfunk-Dienste vor Ort. [27]

8.3.3 Open Mobile Alliance

Allgemein

Die Open Mobile Alliance (OMA) wurde im Juni 2002 von fast 200 Unternehmen gegründet. [28] Im Prinzip ist die gesamte Wertschöpfungskette in dieser weltweiten Allianz integriert. Generell gehören die Mitglieder der OMA einer der folgenden Kategorie an: Mobilfunkanbieter, Applikations- und Contentprovider, Anbieter von drahtlosen Geräten oder IT Unternehmen. In der Zwischenzeit sind noch weitere Unternehmen dazu gestossen, sodass die Allianz jetzt über 350 Mitglieder zählt.

Die Gründung von OMA gestaltete sich als Zusammenschluss der Supporter der Open Mobile Architecture Initiative und derer des WAP Forums. Weitere Foren und Initiativen wurden später integriert, wie die SyncML Initiative, das Location Interoperability Forum (LIF), die Multimedia Messaging Interoperability Group (MMS-IOP), Wireless Village, das Mobile Gaming Interoperability Forum (MGIF) und das Mobile Wireless Internet Forum (MFIW). Die OMA nimmt heutzutage eine zentrale Rolle ein, wenn es darum geht Standardisierungen im Bereich Mobilfunk und generell Wireless zu entwickeln. Das ist auch das primäre Ziel der OMA. Solche Standardisierungen sollen es ermöglichen, „mobile Dienste“ zu kreieren, welche den Bedürfnissen der Endbenutzer entsprechen. Zu den Bedürfnissen gehört auch die Interoperabilität betreffend Länder, Anbieter und Endgeräte. Natürlich sind auch eigennützige Motivationen der jeweiligen Unternehmen entscheidend, zum Beispiel Marktwachstum.

Das Entwickeln von Standards benötigt auch eine gewisse Zusammenarbeit mit anderen Standardisierungsorganisationen. Die OMA arbeitet deshalb mit dem World Wide Web Consortium (W3C), dem 3rd Generation Partnership Project(3GPP), der Internet Engineering Task Force (IETF), dem Java Community Process(JCP), dem European Telecommunications Standards Institute (ETSI), der GSM Association (GSMA), der International Federation of the Phonographic Industry (IFPI), der Recording Industry Association of America (RIAA) und noch weiteren zusammen.

Die übergreifenden Zusammenarbeiten und Bemühungen werden in Abbildung 8.5 verdeutlicht.

Zielsetzungen und Prinzipien der OMA

Zielsetzungen

- Fördern von offenen, qualitativ hoch stehenden Spezifikationen, welche von konkreten Marktbedürfnissen abgeleitet werden. Sie sollen modular und erweiterbar sein, sowie konsistent zu anderen Spezifikationen sein.
- Sicherstellung, dass die Spezifikationen Interoperabilität zwischen verschiedenen Geräten, geografischen Zonen, Dienstanbietern, Mobilfunkanbietern und Netzwerken bieten. Damit sollen auch konkrete Produktimplementierungen betreffend Interoperabilität erleichtert werden.

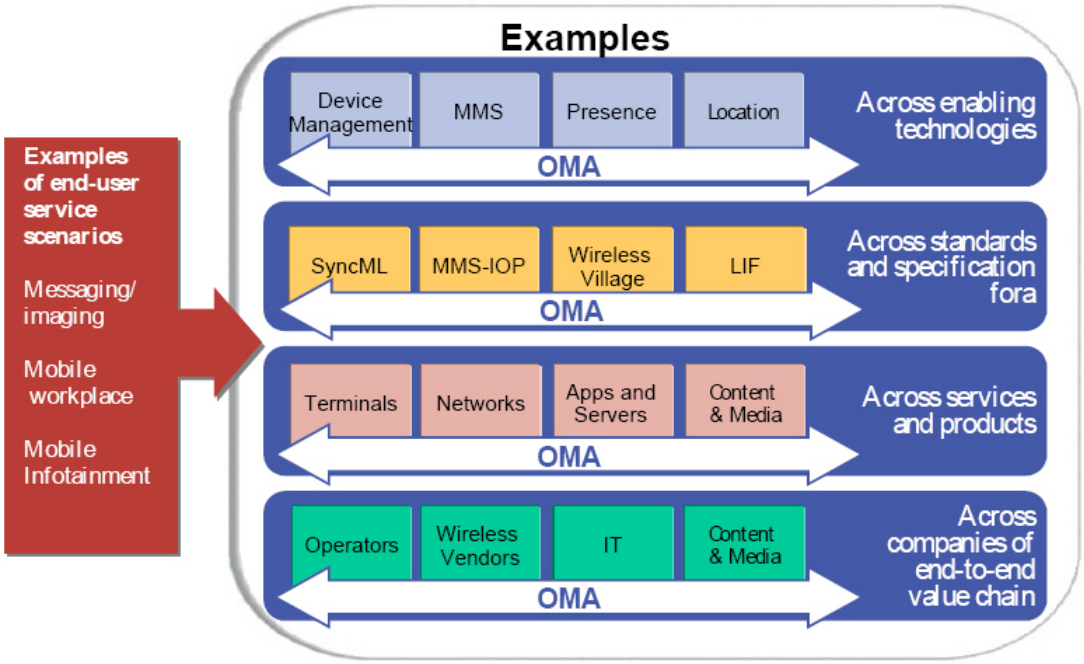

Abbildung 8.5: Verschiedene übergreifende Zusammenarbeiten

- Die OMA soll eine Art Beschleuniger darstellen, um Standards schneller zu konsolidieren. In Zusammenarbeit mit anderen Standardisierungsorganisationen und Industrieforen soll die Interoperabilität erweitert werden und die Kosten aller Beteiligten gesenkt werden.
- Generiert Wert und Nutzen für die einzelnen Mitglieder der OMA, sodass sie auch aktiv daran teilnehmen.

Prinzipien

- Produkte und Dienste basieren auf offenen, globalen Standards. Protokolle und Interfaces werden nicht in proprietäre Technologien eingeschlossen.
- Das Framework und die Service Enablers sind unabhängig vom Betriebssystem.
- Applikationen und Plattformen sind interoperabel und garantieren nahtloses Roaming (orts- und generationsbezogen).

Mehrwertgenerierung durch die OMA

- Für die gesamte Industrie (Mobilfunk)

  Der eigentliche Nutzen resultiert daraus, dass die gesamte Wertschöpfungskette betreffend mobile Dienste und Applikationen gemeinsam and den Spezifikationen und Standards arbeiten. Dadurch erhöht sich die Geschwindigkeit und Effektivität, mit der entwickelt wird bzw. werden kann. Die Investitionen, welche jedes Mitglied tätigen muss, können auf eine Instanz konzentriert werden und müssen nicht auf verschiedene Instanzen verteilt werden.

- Für die einzelnen Kategorien von Mitgliedern

  Lösungen mit offenen Standards generieren immer economies of scale, bei allen Beteiligten der Wertschöpfungskette.

  IT Unternehmen
  Die IT Unternehmen profitieren einerseits von der Verkürzung der Entwicklungszeiten durch die offenen Standards und uniformen APIs. Andererseits stellt die OMA ein Framework zur Verfügung, an welches sie sich halten können. Das führt dazu, dass Produkte, Dienste und Technologien schneller ausgeliefert werden.

  Anbieter von drahtlosen Geräten
  Durch die Standards werden die Entwicklungskosten der Anbieter reduziert. Die Standards haben auch einen positiven Einfluss auf das Wachstum des Marktes für mobile Dienste.

  Mobilfunkanbieter
  Mehrwerte werden durch die gesteigerte Interoperabilität generiert. Zum einen erhöhen die Mobilfunkanbieter dadurch ihre Gewinnmargen, zum anderen können den Kunden transparente Dienste zwischen den verschiedenen Netzwerken angeboten werden.

  Applikations- und Contentprovider
  Durch standardisierte Formate vereinfacht sich die Applikations- und Inhaltsentwicklung und bringt den Firmen finanzielle Vorteile.

Struktur der OMA

Das untenstehende Bild zeigt die Organisation der OMA (Abbildung 8.6). Als übergeordnete, koordinierende und leitende Stelle zeichnet sich das „Board of Directors“ aus. Das „Board of Directors“ vereint Vertreter von verschiedenen Unternehmen, die entweder Sponsor oder Vollmitglied der OMA sind. In der Mitte der Abbildung ist auch das zentrale Element der OMA, das „Technical Plenary.“ Die Sitzung ist verantwortlich für Aktivitäten wie Drafts, Bestätigung und Wartung von Spezifikationen sowie Bereitstellen von Lösungen betreffend der Organisation von OMA. Rechts und links der Sitzung sind zwei Komitees und der untere Teil der Abbildung zeigt die verschiedenen Arbeitsgruppen, die so genannten „OMA Work Groups.“ An den Arbeitsgruppen nehmen auch wieder Vertreter von verschiedenen Unternehmen und Institutionen teil.

Die beiden Komitees werden nun kurz beschrieben. Eine Kurzbeschreibung jeder Arbeitsgruppe ist der Tabelle 8.1 zu entnehmen.

Komitees

Operations and Process Committee:
Dieses Komitee definiert die operationalen Prozesse der technischen Sitzung (technical plenary). Es unterstützt operationale und prozessuale Aktivitäten, macht der technischen Sitzung Prozessvorschläge und erarbeitet weitere unterstützende Massnahmen.

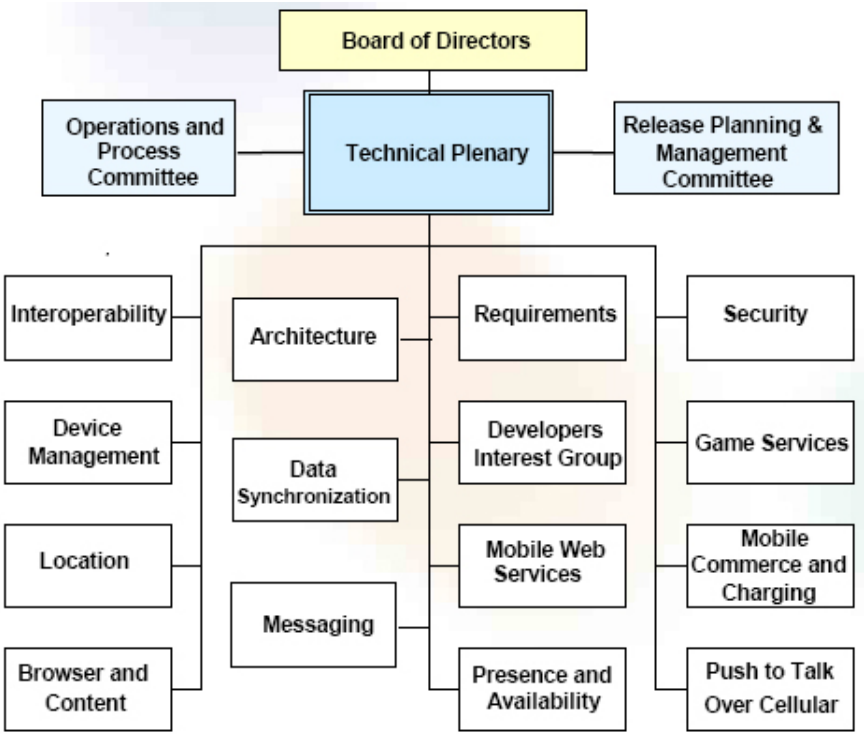

Abbildung 8.6: Übersicht der Struktur der OMA

Release Planning & Management Committee:
Das Komitee ist verantwortlich für die Planung und das Management von OMA Releases. Es definiert die verschiedenen OMA Releases basierend auf den Spezifikationen und kooperiert eng mit den Arbeitsgruppen.

Mitgliedschaft

Um der OMA beizutreten muss die Unternehmung bzw. Institution eine Bewerbung einsenden und nach akzeptierter Aufnahme einen Mitgliedschaftsbeitrag zahlen. Es gibt 4 Arten von Mitgliedschaften: Sponsor, Vollmitglied, Associate und Supporter.

Sponsor: Ein Sponsor hat alle möglichen Rechte und Befugnisse innerhalb der OMA, zusätzlich bekommt er einen Sitz im „Board of Directors." Um Sponsor der OMA zu werden müssen jährlich 140'000 USD bezahlt werden. Beispiele: Mircosoft, Nokia, Orange SA, Vodafone etc.

Vollmitglied: Wie ein Sponsor hat auch ein Vollmitglied alle Rechte und Befugnisse innerhalb der OMA, aber einen Sitz im „Board of Directors" ist nicht sicher. Er hat aber die Möglichkeit, ins Board gewählt zu werden. Ein Vollmitglied zahlt jährlich 35'000 USD. Beispiele: Sony, SonyEricsson, Cisco, VeriSign etc.

Associate: Ein Associate hat nicht mehr alle Befugnisse, er kann beispielsweise nicht im Board vertreten sein, kann nicht in einem der Komitees sein, darf keine WorkGroup leiten, usw. Als Associate der OMA müssen jährlich 7'500 USD bezahlt werden. Beispiel: Swisscom etc.

Tabelle 8.1: Arbeitsgruppen der OMA

Arbeitsgruppe	**Beschreibung**
Architecture	Definiert die gesamte OMA Architektur, welche es den einzelnen Arbeitsgruppen ermöglicht, Spezifikationen zu erarbeiten.
Browser & Content	Diese Arbeitsgruppe ist verantwortlich für die Spezifikationen von Inhaltstypen und deren Darstellung in Browsern. Das soll die Verwendung von datenorientierten Diensten auf mobilen Geräten gewährleisten.
Data Synchronization	Sie führt die Arbeit weiter, welche von der SyncML Initiative begonnen wurde. Vereinfacht gesagt entwirft sie Spezifikationen für die Datensynchronisation.
Developers Interest Group	Sie steht den Entwicklern zur Seite. Das Sammeln und Veröffentlichen von relevanten Daten soll die Entwickler unterstützen. Daneben identifiziert sie auch fehlende und inkonsistente Interfaces, was den Entwicklern weiter unter die Arme greift.
Device Management	Definiert Protokolle und Mechanismen, um die Lebenszyklen der Geräte und Applikationen zu verwalten.
Games Services	Sie definiert Spezifikationen, APIs und Protokolle welche es generell ermöglichen, Spiele zu entwickeln und die Spielentwickler unterstützen sollen.
Interoperability	Identifiziert, spezifiziert und wartet Prozesse und Testprogramme, um Interoperabilität zu gewährleisten. Sie studiert „Best practices“ und versucht diese zu verstehen.
Location	Entwickelt Spezifikationen von Location-Diensten vor allem in Bezug auf Interoperabilität.
Messaging	Ist verantwortlich für verschiedene Messaging-Technologien und definiert auch solche.
Mobile Commerce & Charging	Koordiniert die Anstrengungen der verschiedenen Beteiligten im Feld von M-Commerce.
Mobile Web Service	Sie ist verantwortlich, Spezifikationen betreffend WebService-Applikationen innerhalb der OMA Architektur zu definieren.
Presence & Availability	Presence und Availability Dienste ermöglichen es dynamische Informationen, z.B. Standort, Status über Ressourcen auszutauschen. Diese Arbeitsgruppe entwickelt Spezifikationen welche es ermöglicht diese Dienste einzusetzen.
Push to talk over cellular	Definiert einen Dienst, welcher eine Art Funkgerät-Kommunikation zwischen zwei oder mehr mobilen Geräten ermöglicht.
Requirements	Spezifiziert Use-Cases und erstellet Interoperabilitäts- und Usabilityanforderungen. Die aus den Use-Cases abgeleiteten Anforderungen werden an die Arbeitsgruppen weitergeleitet und zu einem späteren Zeitpunkt wird überprüft, ob die Anforderungen eingehalten wurden.
Security	Erarbeitet Sicherheitsmechanismen für mobile Geräte und unterstützt die anderen Arbeitsgruppen im Bereich Sicherheit.

Supporter: Ein Supporter nimmt an praktisch keiner Aktivität der OMA teil, er darf Dokumente lesen bevor sie der Öffentlichkeit zugänglich gemacht werden, darf Kommentare zu den Release abgeben und darf an den so genannten TestFests teilnehmen. Supporter bezahlen zwischen 250 und 500 USD jährlich.

Vorgehen der OMA

Die OMA hat zum Ziel, aus jeder WorkGroup Spezifikationen zu erstellen, diese sollen dann zu einem späteren Zeitpunkt zu einem Release Enabler werden. Ein Release Enabler ist eine Spezifikation bzw. ein Standard der die Tests der OMA erfolgreich durchlaufen hat und dann konkret implementiert werden kann. Ein genau festgelegter Ablauf wurde entwickelt um aus den einzelnen Spezifikationen ein funktionsfähiges Gebilde zu konstruieren, welches umfassende Interoperabilität gewährleistet. Dieses Release Programm beinhaltet 3 Phasen bzw. Schritte:

1. Phase: Wird auch Candidate Enabler Release genannt. Aus einem Set von Spezifikationen wird ein so genannter Enabler kreiert. Dieser kann in konkrete Produkte und Lösungen implementiert werden, um den Enabler auf Interoperabilität zu testen.

2. Phase: Wird auch Approved Enabler Release genannt. Diese Phase hat ein einzelnes Set von Spezifikationen erreicht und abgeschlossen, wenn die Tests bezüglich Interoperabilität zwischen verschiedenen Geräten und Netzwerken erfolgreich absolviert wurden.

3. Phase: Wird auch OMA Interoperability Release genannt. Ein Enabler hat diese Phase erfolgreich abgeschlossen, wenn die Interoperabilität zwischen verschiedenen Enabler gewährleistet ist. Also im Gegensatz zu Phase 2, wo ein einzelner Enabler auf verschiedenen Geräten und Netzwerken getestet wird, wird nun die Zusammenarbeit mehrerer Enabler auf verschiedenen Geräten und Netzwerken überprüft.

Der Unterschied zwischen Phase 2 und 3 veranschaulicht Abbildung 8.7. Die vertikalen Pfeile beschreiben die Phase 2 und die horizontalen Pfeile die Phase 3.

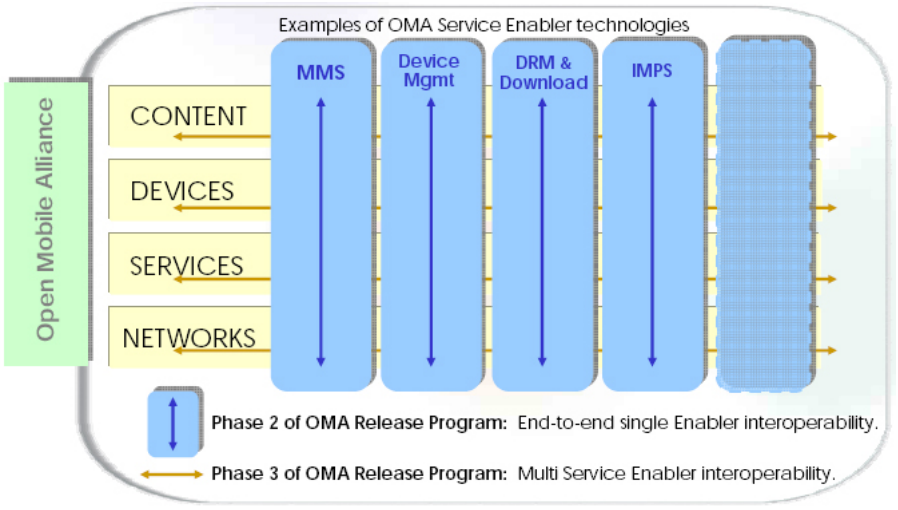

Abbildung 8.7: Unterschied zwischen Phase 2 und 3

Tabelle 8.2: OMA Enabler

Liste aller Enabler (19)	Phase 1: Candiate Enabler Releases	Phase 2: Approved Enabler Releases
OMA Billing Framework	V1.0	
OMA Browsing	V2.2, V2.1	
OMA Client Provisioning	V1.1	
OMA Data Synchronization	V1.2	V1.1.2
OMA Device Management		V1.1.2
OMA Digital Rights Management	V2.0	V1.0
OMA DNS	V1.0	
OMA Download		V1.0
OMA Email Notification	V1.0	
OMA External Functionality Interface	V1.1	
OMA Games Services	V1.0	
OMA Instant Messaging and Presence Service	V1.2	V1.1
OMA Mobile Location Protocol	V3.1	
OMA Multimedia Messaging Service	V1.2	V1.1
OMA Online Certificate Status Protocol Mobile Profile	V1.0	
OMA SyncML Common Specification	V1.2	V1.1.2
OMA User Agent Profile	V2.0, V1.1	
OMA Web Services		V1.0
OMA Wireless Public Key Infrastructure	V1.0	

Die OMA arbeitet derzeit an 19 Enablern. 11 Enabler sind in Phase 1, 8 Enabler in Phase 2 und noch keiner in Phase 3. 10 der 19 Enabler wurden erst im September 2004 zur Verfügung gestellt. Eine Übersicht der Enabler und ihre Versionen sind in Tabelle 8.2 dargestellt.

Ein Enabler wird immer durch mehrere Dokumente beschrieben. Ein „Enabler Release Definition“ Dokument und ein oder mehrere „Specification“ Dokumente werden für jeden Enabler veröffentlicht. Manchmal wird noch ein Dokument betreffend Architektur des Enablers angehängt. Die „Enabler Release Definition“ enthält einen Überblick über den jeweiligen Enabler und zeigt dessen Voraussetzungen für Clients und Server auf. Die Spezifikationen gehen dann detailliert und vor allem technisch auf den Enabler ein. Das Architektur Dokument enthält überwiegend Use Cases und sonstige anwendungsspezifische Details.

Die OMA veranstaltet regelmässig verschiedene Events. Eine Gruppe von Events wird TestFest genannt. TestFests dienen dazu, die Qualität der Spezifikationen zu verifizieren

sowie die Implementationen der einzelnen Unternehmen bezüglich Interoperabilität zu testen. Unternehmen können sich an einem solchen Event treffen und ihre Geräte untereinander testen. Solche Tests sind Voraussetzung, damit die Enabler die Phase 2 abschliessen können. Die Vorteile für die Unternehmen bestehen darin, dass die F&E-Kosten reduziert werden können und eine Früherkennung von Fehlern möglich ist.

Beispiel eines Enablers: Digital Rights Management (DRM)

DRM Systeme wurden entwickelt, um Inhalte und ihre Urheberrechte zu schützen. Sie spezifizieren Befugnisse, Einschränkungen, Verpflichtungen, Konditionen und Übereinstimmungen bezüglich der Nutzung von Inhalten.

Die OMA arbeitete von Beginn ihrer Gründung an den DRM Enablern. Version 1.0 des DRM wurde im November 2002 als Candidate Enabler Release veröffentlicht. DRM 1.0 spezifiziert 3 Methoden, welche sich bezüglich Komplexität und Grad der Sicherheit unterscheiden: Forward Lock, Combined Delivery und Separate Delivery.

Forward Lock erlaubt das Kaufen und Downloaden von Inhalt, aber nicht das anschliessende Weitergeben solcher. Das bedeutet, dass die Inhalte nur auf dem Gerät genutzt werden können, vom welchem sie angefordert wurden. Der Inhalt wird dazu in ein DRM Message gepackt und gegen Weitergabe abgesichert. Abbildung 8.8 veranschaulicht das Prinzip.

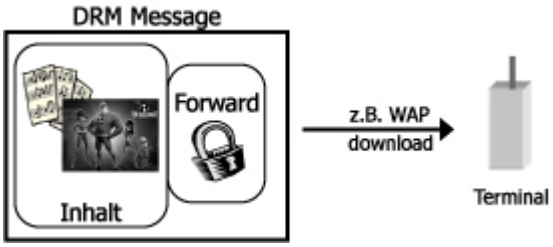

Abbildung 8.8: Forward Lock

Beim *Combined Delivery* wird der DRM Message zusätzlich zum Inhalt noch ein Rights Object hinzugefügt. Das Rights Object dient zur Definition der Rechte über die Nutzung der Inhalte. Als Sprache kommt die Open Digital Rights Language (ODRL) zum Einsatz. Rechte können zeitliche und/oder anzahlmässige Beschränkungen beinhalten. Zum Beispiel kann der Inhalt nur einmal benutzt werden (Anzahl) oder er darf nur eine Woche benutzt werden (Zeit). Die Weitergabe ist auch hier nicht möglich. Die Funktionsweise von Combined Delivery ist in Abbildung 8.9 dargestellt.

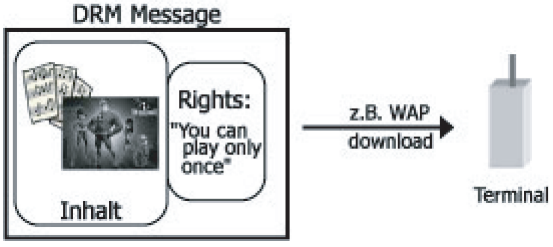

Abbildung 8.9: Combined Delivery

Die letzte Methode *Separate Delivery* funktioniert anders. Das Gerät kann zwar die Inhalte weitergeben, aber die zugehörigen Rechte nicht. Dies wird erreicht durch das separate Verteilen von Inhalten und Rechten über unterschiedliche Kanäle (Abbildung 8.10). Das ermöglicht es Rechte zu erneuern oder abzuändern, ohne die Inhalte neu zu versenden bzw. erneut zu beziehen. Die Inhalte werden durch symmetrische Verschlüsselung ins DRM Content Format (DCF) gepackt. Die Rechte enthalten den Schlüssel zur Entschlüsselung.

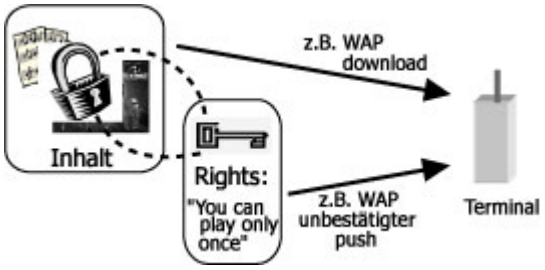

Abbildung 8.10: Separate Delivery

Das DRM 1.0 ist momentan in der Release Phase 2, also ein Approved Enabler Release. Es gibt derzeit schon über 50 Geräte, welche das DRM implementiert haben bzw. kompatibel sind. Jedes Gerät kann einzelne Methoden unterstützen oder auch eine Kombination der Drei.

Im Juli 2004 wurde die Version 2.0 des DRMs veröffentlicht. Version 2.0 baut auf der Version 1.0 auf. Sein Schwerpunkt liegt im Gegensatz zur Version 1.0 auf der Implementierung in mächtigere Geräte, welche auch fähig sind umfangreichere und qualitativ hochwertigere Inhalte zu nutzen. Zusätzlich zu den oben genannten Verfahren in der ersten Version des DRM, enthält die zweite Version zwei wichtige Neuerungen. Um die Sicherheit zu erhöhen wurde ein Public-Key-Verfahren spezifiziert, welches die symmetrischen Schlüssel zur Entschlüsselung der Inhalte schützen soll. Zusätzlich soll die Integrität der Inhalte durch Zertifikate sichergestellt werden. Momentan sind noch keine Geräte mit dem DRM 2.0 kompatibel, deshalb ist es noch ein Candidate Enabler Release.

8.4 Fazit und Ausblick

Die ganze Thematik betreffend Allianzen im Mobilfunkmarkt ist schwierig zu erfassen. Dies hat hauptsächlich zwei Gründe: einerseits ist die Mobilfunkbranche einer rasanten Entwicklung ausgesetzt, welche durch ihre Schnelllebigkeit ständig neue Kooperationsformen fordern. Andererseits ist die noch junge Branche nicht konsolidiert, was viele Neueinsteiger veranlasst in den sich entwickelnden Markt einzutreten.

Bei der Abschätzung der Risiken und Chancen von Allianzen wird ersichtlich, dass Allianzen zweischneidig sind: sie können sowohl konkurrenzfördernd wirken, z.B. durch den Eintritt von Neueinsteigern in ausländische oder branchenfremde Märkte (Reduzierung der Marktkonzentration), als auch den Wettbewerb durch juristisch bedenkliches Verhalten wie z.B. das Errichten von künstlichen Markteintrittsbarrieren durch das Blockieren von wichtigen Ressourcen. Aufgabe der Kartellrechtsbehörden besteht nun darin den Missbrauch zu verhindern.

Das Positive an den heutigen Allianzen ist die Förderung von Technologien, welche durch die Bemühungen der Allianzen in einem neuen Markt schnell zu De-facto Standards werden. Diese ermöglichen den Allianzen ihre Produkte und Dienstleistungen aufeinander abzustimmen und zügig auf dem Markt zu positionieren. Durch die zunehmende Konvergenz von Internet und Telekommunikation entstehen neue Dienstleistungen wie z.B. WAP-Portale oder Streaming-Media-Angebote auf den Handhelds. Technologische Konvergenzen beeinflusst die Entwicklung der heutigen Geräte, vor allem die Smartphones, welche Entertainment, Business-Applikationen und die eigentliche Telefonie miteinander verbinden, sind davon betroffen.

Entscheidend für den Erfolg dieser Innovationen liegt jedoch in der Akzeptanz des Kunden. Die Aufgabe der Allianzen liegt nun darin die neuen Technologien derart einzusetzen dass sie dem Kunden einen Mehrwert bzw. Nutzen bringt. Die heutige Entwicklung hat in gleichem Masse etablierte (SMS) als auch „unausgereifte“ Technologien (WAP) zu verzeichnen. Wie sich dieses Verhältnis in Zukunft entwickeln wird bleibe dahingestellt.

Die Autoren dieser Arbeit sind der Meinung, dass Allianzen langfristig in der globalisierten Welt eine nicht mehr wegzudenkende Marktstruktur darstellen. Die Anforderungen an die Unternehmen in einem sich rasant verändernden und globalisierten Umfeld können nur durch Zusammenarbeit effizient bewältigt werden.

Literaturverzeichnis

[1] Weinhold-Stünzi, Führung, o.A., 1989, S.14

[2] Bischof Köhler zitiert in Kleebach, Strategische Allianzen zur Technologieentwicklung, 1994, S.12

[3] Reuter, Strategische Allianzen und konglomerate Zusammenschlüsse, Verlag Neumann, Jahr, S. 44

[4] Mahrdt Niklas, Strategische Allianzen bei digitalen Informations- und Kommunikationsdiensten, Nomos Wirtschaft Hrsg., 1998

[5] Volkart, Corporate Finance, Grundlagen von Finanzierung und Investition, 2003, S.312

[6] Ordover, Janusz et al., Antitrust Policy and High-Technology Industries, 1998, S.14

[7] Wikipedia, Wikipedia - Standard, [online] 2005, http://de.wikipedia.org/wiki/Standards (Abfragedatum: 07.02.2005)

[8] Wikipedia, Wikipedia - Transaktionskosten, [online] 2005, http://de.wikipedia.org/wiki/Transaktionskosten (Abfragedatum: 07.02.2005)

[9] Schmidt, Wettbewerbspolitik und Kartellrecht, 1993, S.129

[10] Olesch, Strategische Partnerschaften im deutschen und europäischen Kartellrecht., 1992, S.287

[11] Kantzenbach, Erhard et al., Kollektive Marktbeherrschung, Bd. 75,1989, S.8

[12] Lechner, Christoph: Die Entwicklung von Allianzsystemen. Überlegungen an einem Beispiel aus der Telekommunikationsindustrie, Verlag Paul Haupt, 1999.

[13] Jaggi, Stephan G.: Strategische Allianzen im europäischen Telekommunikationssektor. Das Verhaltenskonzept der EG Kommission vor dem Hintergrund U.S. amerikanischer Erfahrungen, Nomos Verlagsgesellschaft, 2003.

[14] Informationszentrum Mobilfunk, Geschichte des Mobilfunks, [online] o.A., http://www.izmf.de/html/de/705.html, 07.02.2005 (Abfragedatum: 07.02.2005)

[15] Die Geschichte von GSM, Die Geschichte von GSM, [online] o.A., http://freeweb.dnet.it/fame/michaelgsm/geschichte.htm (Abfragedatum: 07.02.2005)

[16] AreaMobile, Special: UMTS in Deutschland - Hintergründe und Ausblicke, [online] 2003, http://www.areamobile.de/specials/special_umts_update.php (Abfragedatum: 07.02.2005)

[17] Comcom, Home, [online] o.A., http://www.fedcomcom.ch/comcom/d/homepage/index.html

[18] Münster, P.: Katerstimmung nach dem UMTS-Poker /Fragen nach der Finanzierung der Lizenzen in Deutschland, In: Neue Zürcher Zeitung 19.08.2000, S. 25

[19] Stumpf, Ulrich (et Al): Der Einfluss der offenen Standards auf das Dienstangebot im Mobilfunk der 3. Generation [online] 2004, http://www.bmwa.bund.de/Redaktion/Inhalte/Pdf/der-einfluss-der-offenen-standards-auf-das-dienstangebot-im-mobilfunk-der-dritten-generation,property=pdf.pdf (Abfragedatum: 07.02.2005)

[20] Hautzinger Nina, Siegert Gabriele: Marketing und Vermarktung unter Konvergenz-Bedingungen. In: Karmasin Matthias, Winter Carsten (Hrsg.): Konvergenz und Management. Eine Einführung in die zentralen Herausforderungen und Aufgaben, erscheint noch.

[21] Ecin, AOL und Pillips vereinbaren Zusammenarbeit [online] 2001, http://www.ecin.de/news/2001/07/13/02351 (Abfragedatum: 07.02.2005)

[22] Inside-handy, Mobilfunkunternehmen suchen verzweifelt nach neuen Märkten [online] 2004, http://www.inside-handy.de/news/1527.html (Abfragedatum: 07.02.2005)

[23] Swisscom mobile, Partner: Gemeinsam mehr erreichen [online] o.A., http://www.swisscom-mobile.ch/bus_asp/bus_home.asp?nid=1103&UserLanguage=D (Abfragedatum 07.02.2005)

[24] Vodafone - How are you?, Vodafone - Das Unternehmen, [online] 2005, http://www.vodafone.de (Abfragedatum: 07.02.2005)

[25] starmap mobile alliance, starmap mobile alliance, [online] 2004, http://www.starmapalliance.net (Abfragedatum: 07.02.2005)

[26] T-Mobile, FreeMove Alliance, [online] 2005, http://www.t-mobile.de/freemove (Abfragedatum: 07.02.2005)

[27] Freemove, Freemove, [online] 2004, http://www.freemovealliance.net (Abfragedatum: 07.02.2005)

[28] OMA home, OMA home, [online] 2002, http://www.openmobilealliance.org (Abfragedatum: 07.02.2005)

Chapter 9

Overview and Analysis of Content Provider Business

Christian Müller, Oliver Strebel, Matthias Taugwalder

Der Schwerpunkt dieser Arbeit liegt auf der Analyse verschiedener Modelle zur Klassifikation von Content Providern. Ein erster Teil befasst sich mit den theoretischen Grundlagen: Was sind "'Content Provider"'? Wie lässt sich ein Geschäftsmodell charakterisieren? Welches sind die Merkmale von Geschäftsmodellen internetbasierter Unternehmungen? Anschliessend erfolgt eine Evaluation verschiedener Modelle (u.a. Bambury, Timmers, Rao, Weill/Vitale, Afuah/Tucci und Rappa) und eine Betrachtung der verschiedenen Unterkategorien des Modells von Rappa. Den Schluss bilden eine Anwendung der erhaltenen Klassifikationkriterien auf einige Fallstudien sowie das Ziehen von Schlussfolgerungen.

Inhaltsangabe

9.1 Einführung

Die vorliegende Arbeit soll einen Überblick über eBusiness-Geschäftsmodelle mit Schwerpunkt Content-Provider geben. Nach einer Abgrenzung des Begriffs zu Beginn der Arbeit wird ein Überblick über die bekanntesten und aktuellsten Geschäftsmodelle präsentiert. Eine Gegenüberstellung dieser Klassifikationen rundet dieses Kapitel ab. In einem Hauptteil der Arbeit soll auf das Modell von Rappa eingegangen werden, das einen sehr praxisorientierten Ansatz aufzeigt. Als Abschluss werden die gewonnenen Erkenntnisse in ausgewählten Fallstudien reflektiert.

9.1.1 Abgrenzung

Content Provider werden in den betrachteten Modellen sehr unterschiedlich definiert. Wir gehen von der Content Value Chain aus, die mit der Herstellung des Contents beginnt und nach Veredelung dessen, durch entsprechende Distributionskanäle als Endprodukt zum Konsumenten gelangt, wie in der Abbildung 9.1 ersichtlich ist.

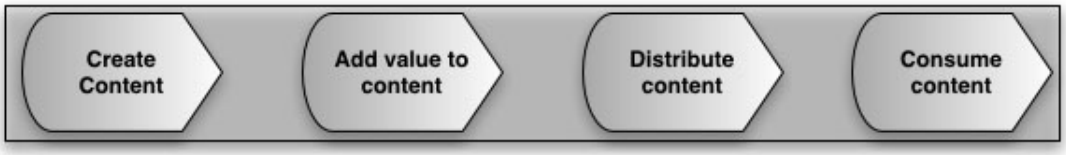

Figure 9.1: Content value chain

Das Wesen und der Umfang eines Content Providers kann sowohl im engeren, wie auch im weiteren Sinne definiert werden (siehe Abbildung 9.2). Ersteres, als engste Definition, orientiert sich am klassischen (News-)Modell und umfasst lediglich Anbieter, die digitale Inhalte in Form von Text, Bild, Ton oder Film produzieren. Diese werden durch Dritte den Konsumenten angeboten und nicht direkt vom Hersteller verkauft. Als klassisches Beispiel dafür wären Journalisten zu nennen, die ihre Artikel an Zeitschriften verkaufen. Eine etwas abgeschwächte Form umfasst zusätzlich auch diejenigen Geschäftsmodelle, die einen eigenen Vertrieb der Inhalte vorsehen, so wie es AOL Time Warner (siehe http://www.aoltimewarner.com) praktiziert (Filmproduktion in Kombination mit eigenen Fernsehsendern). Da sich diese Betrachtungsweisen ausschliesslich auf Inhaltsanbieter beschränkt, die starke Analogien zu herkömmlichen Verlagshäusern und sachverwandten Geschäftsmodellen aufweisen, sind wir der Überzeugung, dass diese Definitionen in jüngster Zeit eher zu kurz greifen.

Wir vertreten die Ansicht, dass beinahe jede Firma, die sich im Internet präsentiert und betätigt, in der einen oder anderen Form als Content Provider auftritt und daher auch gängige eCommerce - Geschäftsmodelle in die Betrachtungen miteingeschlossen werden sollten. Nach Wirtz gehören auch das Zusammenstellen (Packaging) und das Bereitstellen von Inhalten zu einem Content Provider [13]. Insofern ist ein Produktanbieter, der den Kunden das intuitive Vergleichen der Angebote und beispielsweise die Abgabe von

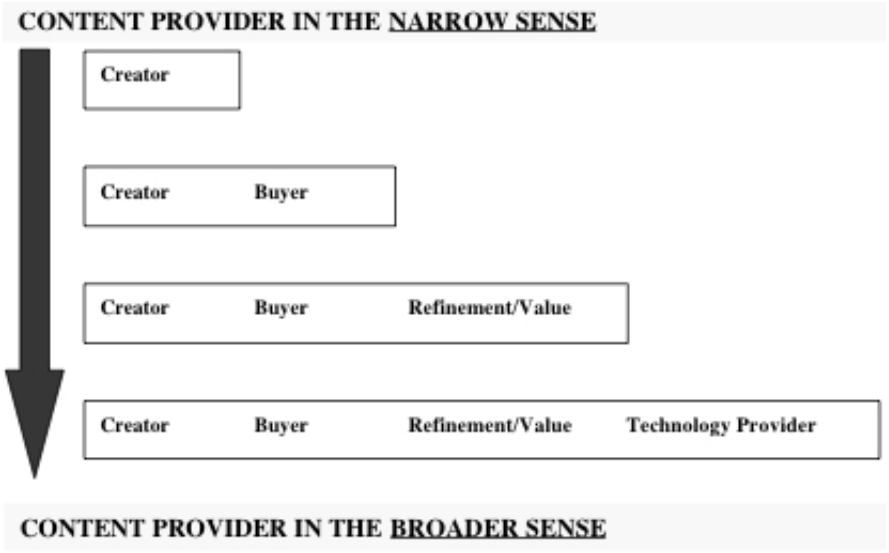

Figure 9.2: Content Provider im engeren und weiteren Sinn

Bewertungen und oder Annotationen ermöglicht, ein Inhaltsanbieter im weiteren Sinne. Diese Form der Inhalte ist im klassischen Handel kaum realisierbar und wenn, dann nicht in diesem Ausmass und dieser Benutzerfreundlichkeit. Insofern bietet das Internet als Plattform einen Mehrwert in Form von zusätzlichem, sinnvollem Content. Demzufolge beziehen wir ein breiteres Spektrum von eBusiness Geschäftsmodellen in unsere Untersuchung mit ein.

9.2 Theoretische Grundlagen

9.2.1 Definition des Begriffs Geschäftsmodell

Wirtz [10] verwendet hierbei einen Strukturierungsansatz mit dessen Hilfe ein Unternehmen, unabhängig von seinem Betätigungsfeld bzw. Unternehmenszweck in Partialmodellen systematisch betrachtet werden kann. Ein Geschäftsmodell ist nach seiner Definition eine Architektur für Produkt-, Service- und Informationsflüsse, einschliesslich der Beschreibung der Akteure mit ihren jeweiligen Rollen, der Gewinne / Leistungen für die verschiedenen Akteure und der Einnahmequellen.

Mit dem Begriff Geschäftsmodell (Business Model) wird hier die Abbildung des betrieblichen Produktions- und Leistungssystems einer Unternehmung bezeichnet. Hierunter fallen das Kapitalmodell (Finanzierungsmodell und Erlösmodell), das Beschaffungsmodell, das Leistungserstellungsmodell, das Distributionsmodell, das Marktmodell (Wettbewerbsmodell und Nachfragermodell) und das Leistungsangebotsmodell als Partialmodelle eines integrierten Geschäftsmodells (siehe Abbildung 9.3).

Nach Wirtz wird ein Geschäftsmodell wie folgt definiert.

"'Das Geschäftsmodell bildet ab, welche externen Ressourcen in die Unternehmung fliessen und wie diese durch den innerbetrieblichen Leistungserstellungsprozess in vermarktungsfähige Informationen, Produkte und/oder Dienstleistungen transformiert werden."' [13]

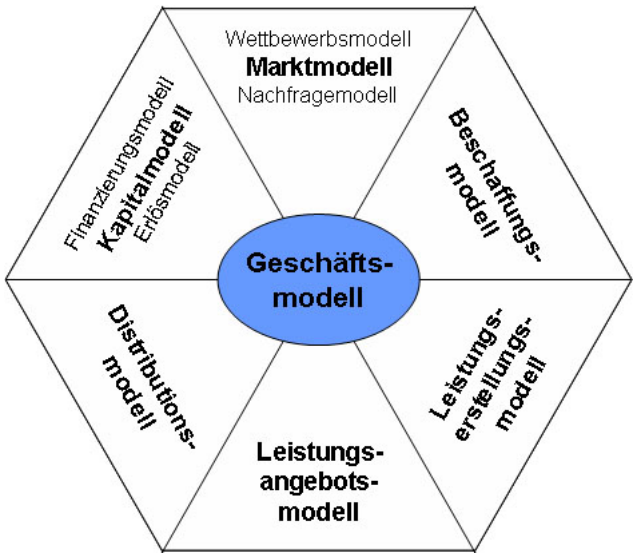

Figure 9.3: Kategorisierung Partialmodelle

Nachfolgend soll das Erlösmodell als Partialmodell herausgegriffen werden, da es für Content Provider ein zentrales Problem adressiert. Generell gliedert sich die Entscheidung wie und in welcher Höhe Erlöse zur Finanzierung der Geschäftstätigkeit erzielt werden soll, in zwei Teilbereiche.

1. Grundsatzenscheidung über die Wahl der Erlöstypen und -Modelle
2. Die konkrete Preispolitik

Bezüglich der Wahl der Erlöstypen (Systematik der Erlösformen, siehe Abbildung 9.4) stellt sich die Frage nach der Art der Verrechnung, ob diese direkt nutzungs- oder transaktionsabhängig oder indirekt über Werbung, Kommissionen oder das Sammeln von Nutzerinformationen erfolgen sollen.

Der Marktplatz für Angebote internetbasierter Unternehmen erstreckt sich theoretisch auf die weltweit angeschlossenen Teilnehmer im Internet, da bislang vorherrschende geographische Grenzen wegfallen. Die wichtigste Erlösquelle bleibt die Werbefinanzierung. Internetwerbung weist steigende Wachstumsraten auf. Dennoch herrscht immer noch eine gewisse Zurückhaltung der Werbetreibenden, was Internetwerbung anbelangt.

Eine weitere Erlösform ist die Kommission, die vor allem bei Partnerprogrammen zum Einsatz kommt. E-Commerce-Anbieter, bislang vorwiegend aus dem Buch- oder Musikbereich, suchen sich reichweitenstarke und zielgruppenaffine Online-Anbieter, auch "'Affiliates"' genannt. Auf deren Seiten im Internet werden Werbungen in Form von Banners etc. geschaltet, die den User auf die Internetsite des E-Commerce-Anbieters weiterleiten.

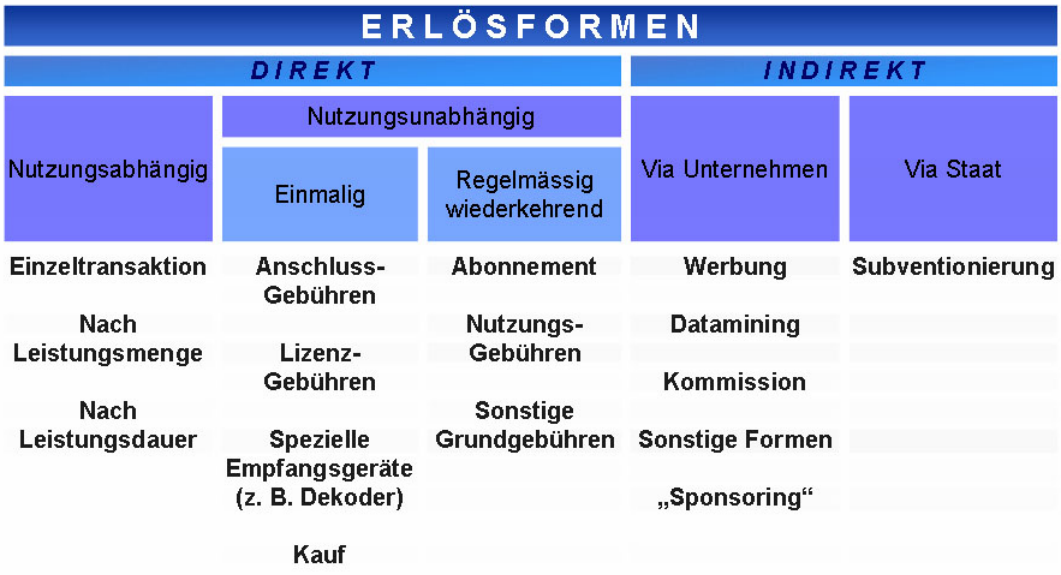

Figure 9.4: Systematik der Erlösformen

Tätigt der "'weitergeleitete"' User einen Kauf, muss der E-Commerce-Anbieter einen vereinbarten Prozentsatz des Kaufumsatzes an den Affiliate rückvergüten. Dadurch, dass das absolute Volumen der Werbeausgaben im Internet im Vergleich mit anderen Medien wie TV oder Zeitschriften immer noch gering ist, bieten Kommissionen eine mögliche Erlösquelle für Online-Medien aus dem Special-Interest-Bereich. Die geringe Reichweite und daraus resultierenden niedrigen Werbeeinnahmen können teilweise durch das Angebot einer spezifischen Zielgruppe und daraus resultierenden Kommissionen kompensiert werden. Für den E-Commerce-Anbieter spielt vor allem der kontextbezogene Inhalt des Online-Publishers zu seinen angebotenen Gütern oder Dienstleistungen eine wichtige Rolle.

Eine Erlösquelle, die für die Zukunft viel verspricht, ist das Datamining. Darunter wird allgemein die Aufzeichnung (Warehousing) und Aufbereitung von Daten und Wissen verstanden, die durch Nutzertransaktionen jedweder Art (Telefongespräche, Fragebögen, Gewinnspiele, etc.) anfallen. Die Refinanzierung erfolgt durch die Weitergabe oder die Bereitstellung der Daten an dritte Unternehmen, die mit den Informationen neue Kunden ansprechen oder ihr Leistungsangebot optimieren möchten. Diese zielgruppengerechte Ansprache der Nachfrager durch die Anbieter wird gerne "'One-to-One"' genannt. Der Vorteil liegt darin, dass sich der Kunde mit keinen unerwünschten Werbeinhalten mehr herumschlagen muss und Botschaften erhält, die gezielt auf seine Bedürfnisse abgestimmt sind. Für den Anbieter ergibt sich eine Minimierung der Streuverluste, wodurch er die vorhandenen Mittel für Werbung gezielter einsetzen kann. Gegenwärtig entspricht One-to-One aber immer noch einer Utopie, deren Erfüllung noch in ferner Zukunft liegt. Der Grund dafür ist die ungenügende Auswertung der Daten. Während die Möglichkeit des Trackings dem Internet Vorteile gegenüber anderen Medien verschafft, lässt sich die anfallende Menge an Daten nur mit grossem Aufwand auswerten.

Tracking ("'Log files"'):

- Nutzung der Hauptseite und der Einzelseiten (Visits, Pageviews) pro Zeiteinheit
- Herkunfts-Domain des Nutzers
- Installiertes Betriebssystem
- Browser
- Werbemittelbezogene Daten (Click-Through etc.)
- Links, die Traffic für die jeweilige Webseite generieren

weiter "'ZAG-Daten"' (Zip Code, Age, Gender):

- Postleitzahl
- Alter
- Geschlecht

Da viele Online-Anbieter sich in der Aufbauphase befinden, können nur ungenügend Ressourcen für die Aufbereitung von Daten zur Verfügung gestellt werden und die aufgezeichneten Daten werden grösstenteils ungenutzt wieder gelöscht.

"'Excite has had enough trouble just keeping up with its phenomenal growth: In February [1998] the company served 28 million pageviews each day, up from 4 million per Day year earlier. At this pace, Excite collects 40 gigabyte of data in ist log files every day. If we tried to look at things at that level, we'd go insane - we'd drown in information"' [3]

Während Abonnementgebühren in den traditionellen Medien wie Printmedien oder Fernsehen zu den wichtigsten Erlösströmen gehören, haben sich Abonnementgebühren im Internet nur teilweise durchsetzen können. Gemäss dem Motto "'follow the free"' zeigen sich User nur zögerlich bereit, für Inhalte aus dem Internet Gebühren zu bezahlen. Dies mag durch die Entstehung des Internets bedingt sein. Die Erlösform des Abonnements eignet sich für zwei Zielgruppen: Erstens für Spezialanbieter von Fachinformationen, deren Nutzer vor allem aus dem Geschäftsbereich stammen. Zweitens können Anbieter auch für Special-Interest-Inhalte Gebühren verlangen, die auf ein hohes Interesse der Nutzer stossen. Beispiele dafür sind Abonnementgebühren für Online-Gaming, Erotic-Channels, oder Sport-News. Generell kann man aber feststellen, dass die Zahlungsbereitschaft der Internetbenutzer stetig zunimmt, auch für allgemeine Inhalte.

Um diese verschiedenen Betrachtungsweisen und Ansätze zusammenzubringen, halten wir uns an die branchenunabhängige Interpretation von Slywotzky:

> "'... A business design is the totality of how a company selects its customers, defines and differentiates its offerings, defines the tasks it will perform itself and those it will outsource, configures its resources, goes to market, creates utility for customers, and captures profit. It is the entire system for delivering utility to customers and earning a profit from the activity. Companies may offer products, they may offer technology, but that offering is embedded in a comprehensive system of activities and relationships that represents the company's business..."' [8]

9.2.2 Merkmale von Geschäftsmodellen internetbasierter Unternehmen

Gerade im Bereich der internetbasierten Unternehmungen ändern sich die Rahmenbedingungen ständig. Im Folgenden möchten wir Eigenschaften von Geschäftsmodellen internetbasierter Unternehmungen noch weiter beschreiben.

Technologieorientierung

Unter Technologieorientierung versteht man die systematische Ausrichtung des Leistungspotentials eines Unternehmens an produkt- und verfahrensbezogener Technik, um dadurch strategische Erfolgspositionen aufzubauen.

Netzwerkorientierung

Unternehmen die sich netzwerkorientiert verhalten, verfügen über die Fähigkeit oder zumindest das Bestreben unternehmensübergreifend erfolgreich entlang der Supply Chain zusammenzuarbeiten. Durch die Kooperation mit anderen Unternehmen und allfälligen Zusammenschlüssen zu so genannten virtuellen Unternehmen, können der Kundschaft neue Produkte mit Qualitäten angeboten werden, die ein einzelnes Unternehmen nicht zu erstellen im Stande gewesen wäre.

Austausch digitaler Güter

Die grosse Verbreitung des Internet hat dem Tauschhandel zu neuer Blüte verholfen. Napster, Gnutella, eMule und Kazaa waren in letzter Zeit in aller Munde. Das Phänomen des Software- und Musiktausches ist nicht erst seit dem Aufkommen des Internets bekannt, entwickelte sich jedoch innert kürzester Zeit in ungeahntem Ausmass. Dass diese Tauschhandlungen nicht nur im Graubereich oder der Illegalität stattfinden, zeigt das Projekt UNIVERSAL [http://nm.wu-wien.ac.at/universal] mit seiner pan-europäischen Brokerage Plattform für Lehrmaterialien. Neben möglichst wirksamen Anreizsystemen die die Bereitstellung von Inhalten fördern sollen, muss auch die Entwicklung von web-gerechten Austauschformaten, Katalogen und die Rechte-Konzeption und Verwaltung geregelt werden.

Nutzung von indirekten Erlösformen

Im Gegensatz zu den typischen realen Kaufgeschäften bei denen der Käufer dem Verkäufer eines Gutes unmittelbar den vereinbarten Kaufpreis bezahlt, sind indirekte Erlösformen im Internet viel stärker verbreitet. Viele digitale Güter werden nach wie vor kostenlos angeboten, da die Kundschaft noch nicht bereit ist, dafür zu bezahlen. Insofern finanzieren sich die Anbieter von kostenlosen Inhalten über Werbeeinnahmen, den Verkauf von Nutzerinformationen oder im Falle einer Auktionsplattform über Transaktionsgebühren.

Trend zu allgemeiner Konvergenz

Die reifen Volkswirtschaften haben sich weltweit in der vergangenen Dekade immer stärker zu Informationsgesellschaften gewandelt. Dieser Trend dürfte sich künftig eher noch verstärken. Ausschlaggebend dafür sind die rasche Diffusion von digitalen Kommunikationstechnologien, Verfahren der Kryptografie und Datenkompression sowie der Aufbau neuer Infrastrukturen mit sehr hohen Übertragungsraten, die über ein einziges Netzwerk den Transport von kombinierten Sprach-, Musik-, Bild-, Video-, und Dateninhalten erlauben. Die rasanten technischen Veränderungen tragen dazu bei, dass bislang getrennte Marktbereiche miteinander verknüpft werden. Dadurch verschwimmen traditionelle sektorale Grenzen und es bilden sich neue Wertschöpfungsketten, die von der Medienverbreitung über Anbieter von Kommunikationsinfrastrukturen und -diensten bis zur Softwareentwicklung reicht. So ist seit längerem ein stetiges Zusammenwachsen von Fernsehen und Internet zu beobachten. Neben dieser technischen Konvergenz ist auch eine inhaltliche Angleichung feststellbar. Wie im realen Leben werden erfolgreiche Konzepte kopiert und das Verhalten der Nutzer gleicht sich an. [12] Insofern verwundert es nicht, dass erfolgreich betriebene Teilbereiche von Internet-Unternehmen nachgeahmt werden und sich somit angleichen.

Dynamik

Das Internet als hochdynamisches Umfeld stellt ganz andere Herausforderungen an Unternehmen als dies in der realen Welt der Fall ist. Die Rahmenbedingungen, meist geprägt durch technische Aspekte, ändern sich ständig und als selbstverständlich betrachtete Werte können innert kurzer Zeit verschwinden und den Einsatz neuer Technologien notwendig machen. Auch das Kundenverhalten ist im Internet-Zeitalter wesentlich unberechenbarer und die Präferenzen der Kundschaft volatiler geworden.

9.3 Evaluation ausgewählter Modelle

Die nachfolgend vorgestellten Modelle sind auf Grund ihrer Aktualität oder wegen ihrer allgemein anerkannten Gültigkeit ausgewählt worden. Hierbei wurde vor allem Wert auf einen klaren Praxisbezug gelegt.

9.3.1 Bambury (1998)

Die Klassifizierung von Paul Bambury [2] fokussiert sich auf den Internet Commerce. Unter diesem Begriff vereint der Autor sämtliche kommerziellen Aktivitäten, die mit dem Internet in Zusammenhang stehen. Man unterscheidet in diesem Zusammenhang zwei Kategorien: Einerseits sollen nachfolgend diejenigen Geschäftsmodelle beleuchtet werden, die bis anhin in der realen Welt zur Anwendung kamen und nun ins Internet transferiert

worden sind (Transplanted real-world Business Models). Andererseits ermöglicht das Internet ganz neue Geschäftsaktivitäten, die erst durch dieses Medium entstanden sind (Native Internet Business Models).

Ins Internet transferierte Geschäftsmodelle

Das Versandmodell Das wahrscheinlich bekannteste und am Besten verstandene Geschäftsmodell ist dasjenige des Versandhandels. Im betrachteten Fall werden die bestehenden Abläufe und Distributionssysteme beibehalten und lediglich ein Webshop als neues Frontend realisiert. Einzig die Werbung und der Verkauf (inklusive Zahlung) werden ins Internet verlagert, um durch einen neuen Markt erschliessen zu können. Die klassische Lagerhaltung und die vorhandene Distributionsinfrastruktur bleiben bestehen und sind weiterhin zentrale Erfolgsfaktoren und entscheiden über den Fortbestand des Unternehmens. Insofern stellt das Internet eine erweiterte Verkaufsplattform dar.

Das Werbemodell Viele Suchmaschinen und frei zugängliche News-Seiten finanzieren sich zumindest teilweise durch Werbung. Was beim Fernsehen die Werbeunterbrechungen sind, sind auf dem Internet die "'Banner-Ads"'. So werden auf stark frequentierten Seiten Werbeeinblendungen platziert und der Auftraggeber bezahlt eine entsprechende Gebühr pro Benutzer, der das Banner anklickt. Das Preismodell orientiert sich an dem der klassischen Fernsehsender. So variieren die Preise je nach Menge der durchschnittlichen Besucher der Seite. Alternativ können auch, wie bereits angedeutet, auch nur die konkreten Klicks auf das Banner in Rechnung gestellt oder die beiden Verrechnungsmethoden kombiniert werden.

Das Abonnements-Modell Dieses Modell wird häufig bei News-Diensten, Informationsdatenbanken und Pornografie-Anbietern eingesetzt. Wie in der traditionellen Gesschäftswelt, wo ein Interessent eine Zeitung abonniert, wird im Internet-Commerce dem Benutzer nach der Bezahlung eines Mitgliederbeitrags (Abonnements) der Zugang zu den Online angebotenen Inhalten freigeschaltet. Diese Inhalte können nicht nur Zeitungsartikel oder Real-Time-News (z.B. http://www.reuters.com) sein. Am Erfolgreichsten angewendet wird dieses Modell von Porno-Anbietern, die Ihren Kunden den Zugang zu nicht jugendfreien Inhalten nur nach Bezahlung einer Abonnementsgebühr ermöglichen. Abonnemente sind sowohl als Tages- Monats-, Jahres- oder Life-Time-Abonnemente gebräuchlich.

Das Gratisproben-Modell Vor allem beim Softwarevertrieb kommt das Gratisprobenmodell sehr oft zur Anwendung. Neben kleinen, unabhängigen Entwicklern, die ihre Produkte gratis zum Download anbieten und die Benutzung für einen beschränkten Zeitraum (typischerweise 30 Tage) oder mit beschränkter Funktionalität zulassen, setzen zurzeit auch vermehrt grosse Softwareschmieden wie Macromedia auf dieses Modell. Der potentielle Benutzer soll durch die kostenlose Testmöglichkeit von der Qualität des Produktes überzeugt werden. Nachdem er sich daran gewöhnt hat, wird er zur Bezahlung des bei

kommerziellen Erzeugnissen meist recht hohen Preises oder des symbolischen Beitrages von einigen US-Dollars im Falle von Shareware (Software von kleinen Entwicklergemeinden) animiert.

Das Direktmarketing-Modell SPAM ist das Resultat der Adaption der klassischen Werbemailings auf das Internet. Auf Grund der fehlenden Kontrolle, Kosten und Beschränkungen, die im realen Leben in den Poststellen und durch physikalische Gegebenheiten erfolgt, ist SPAM zu einem weit verbreiteten Phänomen geworden. Obwohl bereits nahezu alle Empfänger dieser unerwünschten Massenmails eine grosse Abneigung gegen die Urheber entwickelt haben, reisst der Strom der SPAM-Mails nicht ab, da bereits eine Rücklaufquote im Promille-Bereich reiche Umsätze generiert. Der Versand von Werbebotschaften an Millionen von Empfängern innert weniger Stunden wäre in der realen Geschäftswelt undenkbar. Daher hat das Internet den Anwendern des Direktmarketing-Modells ganz neue, ungeahnte Märkte und Benutzergruppen erschlossen.

Das Besitzmodell (Real Estate Model) Bei dieser Ausprägung werden sowohl Web-Space (Platz für das Betreiben einer Homepage), E-Mailadressen und Domainnamen verkauft. Da diese Ressourcen (v.a. Domainnamen und die zugehörigen E-Mailadressen) jeweils einmalig sind, besitzen sie einen gewissen Wert und können gehandelt werden. Da jedoch Firmen und Privatpersonen gegenüber gewissen Eigennamen geistiges Eigentumsrecht geltend machen, muss ein nach dem Besitzmodell operierendes Unternehmen genau abwägen, welche Domainnamen es in Besitz nehmen will und kann, ohne die Rechte Dritter zu verletzen.

Das Anreizmodell Die Ausprägungen dieses Modells sind so vielfältig wie die Anreize auf welche potentielle Kunden reagieren oder zu akzeptieren bereit sind. So ermöglicht beispielsweise die Akzeptanz der Einblendung von Werbung die kostenlose Nutzung eines Browsers (im Falle von Opera) oder der Download einer Software ist erst nach dem Ausfüllen einer Umfrage möglich. Das weit verbreitete P2P (Peer to Peer) - Programm Kazaa andererseits funktioniert nur in Kombination mit ebenfalls zu installierenden Data-Miner, die das Surfverhalten des Benutzers protokollieren und teilweise sogar zu beeinflussen versuchen.

Native Internet Geschäftsmodelle

Entgegen der allgemein vorherrschenden Meinung werden die meisten Transaktionen im Internet abgewickelt, ohne den physischen Transfer von Zahlungsmitteln. Viele Produkte und Dienstleistungen werden auch heute noch ohne die Bezahlung eines konkreten Preises erbracht. Die Erzeuger versprechen sich davon die Akkumulation von Reputation. Der Internet-Markt kann im Gegensatz zur durch Knappheit der Ressourcen geprägten realen Wirtschaft durch einen Überfluss an Diensten und Informationen charakterisiert werden.

Bibliothek Das Internet und das World Wide Web im Besonderen sind wahre Informations-Goldgruben. Vor allem Wissenschafter und Akademiker haben sich schon früh daran gemacht, diese Informationen zu sammeln und aufzubereiten, so dass andere daran teilhaben können. Dies geschieht für den Leser meist kostenlos und verbessert die Reputation des Anbieters. Viele Seiten funktionieren auch heute noch nach diesem Prinzip.

Freeware Das Internet in seinem heutigen Umfang wäre ohne Freeware undenkbar. Viele grundlegende Techniken (BIND, Sendmail, Apache, PERL, etc.) werden kostenlos als Freeware vertrieben und ermöglichen so das Wachstum und die schnelle Ausbreitung des Internets. Netscape, der einstmals dominante Internetbrowser wurde vorerst auch als Freeware angeboten und sobald seine Verwendung allgegenwärtig war, wurden erweiterte Versionen gegen Entgelt verkauft.

Tauschhandel Der Tauschhandel ist die gängigste Geschäftsform im Internet. Vielerorts werden digitale Produkte im Austausch gegen persönliche Benutzerinformationen angeboten. Dies kann in Form eines Fragebogens, durch die Installation von Data-Miner-Software oder anderweitig erfolgen. Diese Form des Handels kommt ohne finanzielle Transaktionen aus. Vor allem News-Dienste funktionieren nach diesem Modell.

Digitale Produkte & Lieferung Die Tatsache dass digitale Produkte nicht materiell existieren müssen (aber können!), stellt die beteiligten Parteien vor neue Herausforderungen. So muss beispielsweise im Falle von Informationen die Lieferung zeitlich nach der Bezahlung erfolgen, da sobald die Information bekannt gegeben worden ist, ihr Wert gegen Null tendiert. Unter digitalen Produkten werden neben Informationen, Software, Bild & Ton, auch Zertifikate verstanden.

Zugangsversorgung Der garantierte Zugang zum Internet ist für die Existenz des Internets fundamental. Ohne Benutzer ist die Funktion des Internets ausgeschlossen. Diese Dienstleistung der Internet Service Provider (ISP) wird in der Internet-Commerce Diskussion häufig ignoriert. Die ISPs stellen die Infrastruktur zur Verfügung die notwendig für den Anschluss der Benutzer (Firmen, Private) und das Routing und Einspeisen deren Datenpakete in das weltweite Netz ist.

Hosting und artverwandte Dienste Viele ISPs aber auch reine Hosting-Provider bieten Web-Space (Fesplattenplatz für Homepages), E-Mailadressen und andere verwandte Dienste an. Diese werden entweder durch ein jährliches Abonnement abgegolten oder durch Werbeeinblendungen finanziert.

9.3.2 Timmers (1998)

Paul Timmers zeigt auf, dass zu einem Geschäftsmodell Produkt-, Dienstleistungs- und Informationsflüsse gehören. Diese werden von einem Akteur zu einem anderen gesendet.

Zusätzlich werden auch die dabei erwarteten Vorteile bzw. Umsatzquellen als Teil des Modells definiert. Mit seinem 1998 veröffentlichten Artikel "'Business Models for Electronic Markets"' legte er den Grundstein für die Diskussion um die Internet-Geschäftsmodelle [10]. Neben den bereits erwähnten Ebenen der Beschreibung der Geschäftsmodelle verfolgt er zusätzlich einen systematischen Analyseansatz zur Ableitung von Geschäftsmodellen. Dieser bezieht sich auf die Wertschöpfungsketten und deren De- und Rekomposition, wobei er dem Ansatz von Porter folgt. Die Zerlegung der Wertschöpfungskette erfolgt in drei Stufen. Zuerst wird sie in die Funktionen (nach Porter) aufgespalten [6], danach werden die Interaktionsmuster bestimmt und in die Kategorien 1 to 1, 1 to n, n to 1 eingeteilt. Abschliessend wird die Wertschöpfungskette wieder zusammengebaut. Somit lässt sich der Aufbau eines Unternehmens sowohl nach innen wie auch nach aussen abbilden. Timmers unterscheidet 11 Geschäftsmodelle.

e-Shop

Durch die Präsenz im Internet und durch den Einsatz eines e-Shops eröffnen sich dem Unternehmen neue Möglichkeiten. Neben der stark vergrösserten Reichweite, kann den potentiellen Kunden auch eine 24-Stundenverfügbarkeit, ein neues Einkaufserlebnis, Annehmlichkeiten in Bezug auf Produktvergleiche und Online-Bezahlung angeboten werden. Insofern bietet sich der e-Shop neben den Werbemöglichkeit als ernstzunehmende Verkaufsplattform an.

e-Procurement

Die Online-Beschaffung ist vor allem im Business to Business Bereich interessant. Da die Produktkataloge, Offerten und auch die Verhandlungen im Internet geführt werden, können signifikant Kosten eingespart und durch das Zusammenbringen von vielen Anbietern und Nachfragern die Preise tief gehalten werden.

e-Auction

Elektronische Auktionen stellen das Analogon zu den traditionellen Auktionen dar. Sie bieten sich vor allem für Verkäufer an, die überschüssige Lagerbestände abbauen wollen, und für Käufer, die günstig Waren beziehen möchten, an. Der Anbieter der Plattform verrechnet meist transaktionsbezogene Gebühren um seine Tätigkeit zu finanzieren.

e-Mall

Die elektronische Mall zeichnet sich durch eine Kollektion von e-Shops, meist verbunden mit einem einheitlichen Zahlungssystem, aus. Den einzelnen Shop-Anbietern wird eine einheitliche Plattform zur Verfügung gestellt, was die Unterhaltskosten senkt und gleichzeitig potentiellen Kunden die Möglichkeit eröffnet, an einem Ort (einer konkreten

URL) bei verschiedene Anbieter einzukaufen. Als Beispiel wäre die Electronic Mall Bodensee (http://www.emb.ch) anzuführen, die verschiedenste Unternehmen und Informationen aus der Region Bodensee unter einem Dach vereint.

3rd Party Marketplace

Bei diesem Geschäftsmodell wir das Web-Marketing einer dritten Partei überlassen und die Angebote lediglich in die bestehenden e-Shops integriert, um das Produktportfolio abzurunden. Auch die Bezahlungsmodalitäten gegenüber dem Endkunden können somit bei Bedarf vom Marktplatz bereitgestellt werden.

Virtual Communities

Virtuelle Gemeinschaften leben von den Beiträgen ihrer Mitglieder. Im kommerziellen Bereich tauchen sie meistens als Ergänzung zu bereits bestehenden Marketinganstrengungen auf, um die Kundenbindung und die Attraktivität bestehender Inhalte zu vergrössern. Ein Beispiel bietet der Buchandel mit Amazon (http://www.amazon.com), wo sich verschiedene Benutzer über bereits gekaufte oder geplante Anschaffungen von Büchern austauschen können.

Value Chain Service Provider

Diese Anbieter fokussieren sich auf ihre Position in der Wertschöpfungskette und bieten spezialisierte Produkte wie beispielsweise den elektronischen Zahlungsverkehr oder Logistikdienstleistungen (Online-Tracking, etc.) mit Hilfe von Netzwerken im Allgemeinen und dem Internet im Speziellen an.

Value Chain Integrator

Integratoren der Wertschöpfungskette versuchen den Informationsfluss zwischen den verschiedenen Akteuren der Wertschöpfungskette zu optimieren und somit Ertrag zu generieren. Dies kann durch effizientere Kommunikationsformen und - Techniken oder durch Anreicherung bestehender Informationen erreicht werden.

Collaboration Platforms

Virtuelle Kollaborationsplattformen werden vor allem im Business to Business Bereich zur Unterstützung von virtuellen und verteilten Teams oder im Engineering eingesetzt. Erlöse werden entweder durch einmalige Gebühren oder Abonnements generiert. Als Beispiel wäre Groove Networks (http://www.groove.com) zu nennen.

Information Brokers

Da das Internet vor allem aus einer Ansammlung unzähliger Informationen besteht, verwundert es nicht, dass ein breites Feld von Informations-Maklern entstanden ist. Inhalte im weitesten Sinne werden entweder kostenlos oder gegen Entgelt (Abonemment, Pay per view) angeboten oder gegen andere Güter getauscht.

Trust Services

Die Trust Services stellen eine spezielle Kategorie dar, die meist von Zertifikationsauthoritäten oder elektronischen Notaren angeboten werden. Diese speziell akkreditierten Unternehmen stellen beglaubigte Zertifikate und artverwandte Inhalte zur Verfügung, um beispielsweise eine zweifelsfreie Identifikation von elektronischen Handelspartnern zu ermöglichen.

9.3.3 Rao (1999)

Die Klassifikation von Bharat Rao orientiert sich stark an den an einer Transaktion beteiligten Parteien. So werden die Modelle anhand der Akteure in Business und Consumer Interaktionen aufgeteilt, wobei der Staat (Government) als möglicher Teilnehmer nicht behandelt wird. [7]

Business to Consumer

Der Business to Consumer Bereich stellt den klassischen Markt dar. Rao konzentriert sich in seinen Ausführungen auf dieses Segement und unterteilt ihn weiter in Teilbereiche, wie nachfolgend aufgezeigt werden soll.

Käufer-Seite Käuferseitige Händler versuchen mit Hilfe des Internets grosse Nachfragermassen zu bündeln und dadurch Mengenrabatte bei den Lieferanten der gewünschten Güter auszuhandeln. Sie stellen eine Plattform zur Verfügung, auf der sich Nachfrager in Gruppen zusammenschliessen können. Das Kapital dieser Anbieter liegt in der angebotenen Platform, dem Lieferantennetzwerk und der Möglichkeit einmal gewonnene Kunden möglichst stark zu binden.

Verkäufer-Seite Verkaufsseitige Anbieter entsprechen dem klassischen Einzelhandel, erweitert um eine neue Verkaufsplattform. Beispielsweise Amazon.com kann als erweiterter Buchladen verstanden werden, dessen Ziel es ist, ein möglichst umfassendes Angebot einer weltweiten Kundschaft verfügbar zu machen.

Logistik & Infrastruktur Einige Unternehmen suchen Ihre Chancen darin, die traditionell ineffizienten Beschaffungsketten (Supply Chain) durch den Einsatz von in die physische Distributionsinfrastruktur integrierter IT und der Anbindung ans Internet zu optimieren. Ein wenig erfolgreiches Beispiel liefert Webvan [9], das die Nahrungsmitteldistribution revolutionieren wollte und im Jahr 2001 die Tore schliessen musste.

Business to Business

Der Business to Business Markt ist für Bharat Rao von zentraler Bedeutung. Er vertritt die heute noch gültige Meinung, dass B2B der eigentliche Motor des Internets ist und vor allem sogenannte Metamediäre die Unternehmen aus verschiedenen Marktsegmenten und mit unterschiedlichen Produkten zusammenbringen und somit neue Märkte erschliessen.

Consumer to Consumer

Wie der Begriff suggeriert, interagieren in diesem Segment ausschliesslich die Konsumenten untereinander. Die bekannteste Ausprägung findet man bei Online-Auktionen wie eBay. Hier treten Konsumenten sowohl als Anbieter als auch als Nachfrager auf und greifen lediglich auf eine bereitgestellte Plattform zurück.

9.3.4 Weill / Vitale (2001)

Peter Weill und Michale R. Vitale gehen von atomaren eBusiness-Modellen aus [11], die jeweils dieselben Komponenten beinhalten. So müssen alle Geschäftsmodelle unter anderem über die notwendige Infrastruktur verfügen, strategische Ziele definieren und neben den eigenen Kernkompetenzen auch über eine nachhaltige Einnahmequelle verfügen. Verlässt man die Ebene der Geschäftsmodelle und ergänzt sie durch Implementationsmassnahmen, Erlösmodelle und Marketing, so lassen sich diese auch noch zusätzlich nach Kundensegmenten und den bereits genannten Kriterien unterteilen (Taxonomie dargestellt in Abbildung 9.5).

Direct to Customer

Das traditionelle Retailgeschäft war relativ einfach aufgebaut. Die Händler kaufen Produkte, lieferten diese in die Verkaufslokalitäten und verkauften sie an den Endkunden. Das eBusiness Direct to Customer Modell ermöglicht es den Einzelhändlern, eine riesige Auswahl an Produkten anzubieten. Dies wäre im traditionellen Ladengeschäft nicht möglich. Die Möglichkeit Zahlungen auch online abzuwickeln, Aufträge entgegenzunehmen und die Lieferung direkt auslösen zu können, senkt die Kosten für den Anbieter massiv. Die Haupteinnahmequelle besteht im Verkauf der Waren an die Endkunden. Daneben können jedoch auch durch den Verkauf von Nutzerinformationen und Gebühren für Product Placement oder Werbung zusätzliche Gewinne realisiert werden. Die Unternehmung

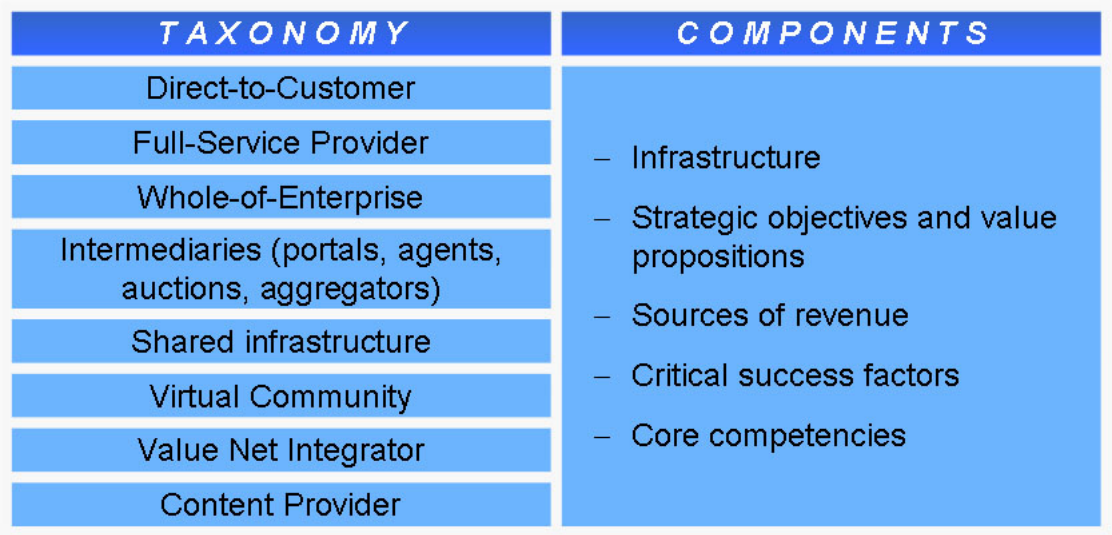

Figure 9.5: Taxonomie nach Weill/Vitale

besitzt bei diesem Modell sowohl die Kundendaten als auch die Kontrolle über die Kundenbeziehung und Kundentransaktionen. Für den Kunden eröffnet sich neben dem stark vergrösserten Angebot auch die Möglichkeit, schnell und einfach verschiedenste, auch geographisch weit verteilte Anbieter miteinander zu vergleichen und somit seine eigenen Transaktionskosten (Such- & Vergleichskosten) zu senken.

Full-Service Provider

Der Full-Service Provider versucht die gesamten Bedürfnisse des angepeilten Kundensegments zu befriedigen. Dies geschieht sowohl durch eigene Angebote, als auch durch Produkte und Dienstleistungen von ausgewählten Drittfirmen, die unter dem Namen des Full-Service Providers an den Endkunden vermittelt werden. Die Partnerfirmen sind bereit für die erhofften Mehrumsätze auf den Kundenkontakt zu verzichten. Der Full-Service Provider auf der anderen Seite rückt näher an den Kunden heran, lernt seine Bedürfnisse kennen und kann sich daher proaktiv verhalten und potentielle Verkaufschancen ausnutzen anstatt bloss auf Kundenanfragen zu reagieren. Als kritische Erfolgsfaktoren wären in diesem Modell das Customer Relationship Management und das Customer Information Management zu nennen.

Whole of Enterprise

Dieses Modell verfolgt ein ähnliches Ziel wie das Full-Service Provider Modell. Dem Kunden soll ein möglichst umfassendes Produkt- und Dienstleistungsportfolio angeboten werden. Beim Whole of Enterprise Modell stammen alle diese Angebote aus demselben Konzern und werden lediglich über einen einzigen "'Point of Contact"' (Interaktionsschnittstelle zwischen Kunde und Konzern) an den Kunden gebracht. Dies kann beispielsweise eine Website sein. Dadurch wird dem potentiellen Nachfrager die Navigation durch die verschiedenen Angebote der Unternehmung erleichtert und er findet sich besser zurecht.

Portals, Agents, Auction Aggregators

Die im Titel genannten Zwischenhändler oder Vermittler bringen Anbieter und Nachfrager zusammen und belasten entsprechende Provisionen, (z.B. abhängig vom Transaktionsvolumen oder durch wiederkehrende Listing-Gebühren) um Erlöse zu generieren. Diese Intermediäre senken die Such- und Transaktionskosten der beteiligten Parteien und sind deshalb sowohl im eBusiness wie auch im traditionellen Handel präsent.

Shared Infrastructure

Teure Infrastruktur-Investitionen werden häufig auf mehrere Partner verteilt um effektiver im Markt bestehen zu können. Orange beispielsweise nutzt Teile der Swisscom - Infrastruktur (Mobilfunkantennen). Durch dieses Modell können einerseits Eintrittsbarrieren für neue Konkurrenten aufgebaut und andererseits Kosten durch Economies of Scale gesenkt werden. Erlöse werden meist durch Mitgliederbeiträge realisiert. Das angesprochene Modell kann jedoch nur funktionieren, wenn keiner der Beteiligten zu dominant wird und eine kritische Masse an Mitgliedern geworben werden kann.

Virtual Community

Virtuelle Gemeinschaften ermöglichen es den Mitgliedern mit Gleichgesinnten zu interagieren und sich auszutauschen. Die Anbieter solcher Plattformen finanzieren sich durch Werbeeinnahmen, den Verkauf von Gütern oder durch Mitgliederbeiträge. Solche Plattformen stehen und fallen mit der Anzahl und Loyalität ihrer Mitglieder.

Value Net Integrator

Durch das Internet wurden die traditionellen Wertschöpfungsketten aufgebrochen und gut positionierte Unternehmen beginnen vermehrt ihr Wissen über die verschiedenen Akteure in diesen Wertschöpfungsketten auszunutzen, um bestehende Abläufe zu optimieren. Durch diese Optimierung entstehen Wertschöpfungsnetzwerke, die flexibel an neue Kundenbedürfnisse angepasst werden können. Einnahmen entstehen in diesen Netzwerken meist durch Mitgliederbeiträge oder Abgaben auf die Produkte, die durch das Netzwerk geschleust werden.

Content Provider

In unserer Arbeit fokussieren wir auf die Content Provider. Weill und Vitale fassen den Begriff des Content Providers ziemlich eng. Sie definieren ihn als Produzent von Inhalten (Text, Bild, Ton, Film) die über einen Dritten meist kostenlos dem Konsumenten zur Verfügung gestellt werden. Die Einkommensquelle sehen sie in erster Linie in den Gebühren die der Content-Hersteller den Partnerfirmen verrechnet. Daneben räumen sie

aber auch ein, dass gewisse Content Provider sowohl die Informationen herstellen und auch gleich selber den Konsumenten zur Verfügung stellen und sich durch Mitgliederbeiträge, Pay-Per-View oder ähnlichen Modellen finanzieren. AOL TimeWarner wäre ein solches Beispiel. In dieser Betrachtung fehlt die Analyse des Wettbewerbsumfelds und bereits vorhandener oder möglicher Wertschöpfung.

9.3.5 Afuah / Tucci (2001) und Rappa (2002)

Afuah und Tucci definieren ein Internet Geschäftsmodell als die Methode mit der eine Unternehmung durch Benutzung des Internets langfristig Erlöse erwirtschaftet. [1] Rappa seinerseits definiert ein Business-Modell danach wo ein Unternehmen in der Wertschöpfungskette steht und auf welche Art und Weise es Einkommen erwirtschaftet, was wiederum nur durch die eindeutige Identifikation seiner Position in der Wertschöpfungskette erfolgen kann. [22]

9.3.6 Gegenüberstellung

Als Rahmen für eine Analyse der verschiedenen Theorien wurde das Modell von Rao gewählt. Seine Unterteilung ist sehr allgemein gehalten und unterscheidet nur nach den verschiedenen Akteuren in einem Geschäftsnetzwerk. Geschäftskunden (Business) und Endkunden (Consumer) und deren Beziehungen untereinander bilden den Fokus der Betrachtung von Rao.
Timmers und Rappa auf der anderen Seite führen sehr detaillierte Modelle ins Feld. So umfasst beispielsweise die Klassifikation von Rappa neun verschiedene Geschäftsmodelle mit gegen 40 Submodellen. Weill / Vitale ihrerseits unterteilen die Ausprägungen ihrer Taxonomie noch weiter in verschiedene Komponenten die sie für das Gelingen des entsprechenden Geschäftsmodells als wesentlich erachten.
Wir haben in der nachfolgenden Abbildung 9.6 alle in dieser Arbeit betrachteten Modelle unter dem Dach von Rao vereint. Dies ermöglicht eine bessere Beurteilung derselben.

Überschneidungen Die verschiedenen Klassifizierungen sind sich in grossen Teilen sehr ähnlich. Beispielsweise werden virtuelle Gemeinschaften von allen Autoren in ihr Schema aufgenommen. Andererseits tauchen ähnliche Geschäftsmodelle bei mehreren Autoren unter anderen Benennungen auf. So führt etwa Rappa den Begriff des "'Infomediary"' und Timmers den des "'Information Brokers"' auf. Insofern ist der Schluss zulässig, dass sich alle betrachteten Schemata derselben Vorgehensweise bedienen und sich in erster Linie in ihrem Umfang unterscheiden.

Gewichtung Bambury legt einen sehr starken Fokus auf den Business to Consumer Markt. Beinahe alle seiner Modelle sind in diesem Teil angesiedelt. Der Consumer to Consumer Markt wird auch von den anderen Autoren nur mit verhältnismässig wenigen Modellen bedacht. Hier zeigt sich, dass Rappa und Timmers durch eine ausgewogene Verteilung auffallen. Rappa kann (unter Einbezug der 40 Submodelle) als umfassendstes

Figure 9.6: Vergleich der verschiedenen Klassifizierungsmodelle

Klassifikationsschema betrachtet werden. Da es seinerseits zu einem guten Teil auf Timmers aufbaut, soll das Modell von Rappa in den nachfolgenden Kapiteln vertieft betrachtet werden und auch als Grundlage für die Fallstudien dienen.

9.4 Modell von Rappa

Nach Hedmann und Kalling [23] lassen sich Geschäftsmodelle in zwei wesentliche Richtungen einordnen: Eine Richtung befasst sich mit generischen Beschreibungsansätzen, die zweite Richtung mit einzelnen, praxisorientierten und ausgewählten Geschäftsmodellen. Die Klassifizierung von Michael Rappa [22] beschreibt Geschäftsmodelle, die sich eher der zweiten Richtung zuordnen lassen. Michael Rappas Klassifikation nennt neun verschiedene Grundmodelle, die für eine präzisere Beschreibung in weitere 40 Submodelle gegliedert werden.

Aufgrund der praxisnahen Modelle haben wir uns dazu entschieden, unsere Übersicht über das Content Provider Business anhand der Klassifikation von Michael Rappa zu gestal-

ten. Unserer Meinung nach hat Michael Rappa - Professor für Informations Technology am College of Management der North Carolina State University und ehemaliger Professor am M.I.T. - einen interessanten Grundstein gelegt, der für weiterführende Arbeiten aufgegriffen und weitergeführt werden kann und soll.

Die einzelnen Modelle werden durch textuelle Beschreibungen und anhand von konkreten Beispielen erläutert, grafische Elemente wie Flussdiagramme oder ähnliches fehlen.
Die Klassifizierung ist - wie Rappa ausdrücklich angibt - nicht erschöpfend, es können also nicht alle Bereiche des e-Commerce durch Modelle abgedeckt werden.

> "'The proposed taxonomy is not meant to be exhaustive or definitive. Internet business models continue to evolve."' [22]

Auf eine strikte Ein- und Abgrenzung der Modelle und auf eine präzise Definition wird bewusst verzichtet. Der Vorteil dieser Vorgehensweise liegt in der Möglichkeit, neue Geschäftsmodelle in die Klassifikation zu integrieren. Damit wird dem Umstand Rechnung getragen, dass sich Geschäftsmodelle stets weiterentwickeln und dass eine klare Abgrenzung der Modelle nicht nur sehr schwierig ist, sondern auch zu einem Verlust der Aussagekraft führen kann. Das praxisorientierte Klassifikationsschema zeigt ja gerade, dass viele Modelle nur in hybrider Form existieren können.

Ein grosser Nachteil dieser Vorgehensweise zeigt sich aber darin, dass sich die Modelle nur schwer miteinander vergleichen und bewerten lassen. Für eine weiterführende Analyse der verschiedenen Geschäftsmodelle bietet die Klassifikation von Michael Rappa nur wenige Kriterien. Für eine Analyse müssen also aussagekräftige Variablen ausgewählt werden, welche die Akteure im Geschäftsnetzwerk samt ihrer Rollen, Austauschbeziehungen und Nutzenpotentiale beschreiben können.[4]

Weil uns bei Geschäftsvorgängen die Einteilung der Akteure in einem ersten Schritt grundlegend erscheint, haben wir die Einteilung der Akteure in den Business-und Consumer-Bereich gemäss Rao, in unsere Übersicht der Modelle miteinbezogen. (vgl. Kapitel 1.3.4)

Als ein weiteres Kriterium soll eine klare Einteilung der Erlöse für angebotene Güter und Dienstleistungen dienen. Dafür orientieren wir uns an der Systematik der Erlösformen von A. Zerdick et al., die wir für unsere Bedürfnisse leicht modifiziert haben. [14]

Im folgenden werden nun die neun Grundmodelle (basic categories) nach Michael Rappa kurz vorgestellt und erläutert.

9.4.1 Brokerage Model

Definition: Beim Brokerage Model erzielt der Betreiber seine Erlöse durch die Vermittlung von Transaktionen zwischen einem Anbieter und einem Nachfrager.

Erlösform: indirekt, via Unternehmen/Kunde, Transaktionskommission

Ziel eines Brokers ist es, auf einer Plattform Anbieter von Gütern und Dienstleistungen mit Kunden zusammenzubringen, wodurch Kaufgeschäfte entstehen. Das Brokerage Model kann in weitere Subformen unterteilt werden. Die wohl bekanntesten Subformen sind die Auktionsbroker (auction broker), wie z.B. Ebay oder Ricardo. Diese Online-Auktionshäuser übernehmen bei der Geschäftsabwicklung keine aktive Rolle. Der Vertrag kommt zwischen dem Einlieferer/Verkäufer eines Guts und dem Ersteigerer/Bieter zustande. Das Online-Autkionshaus stellt lediglich eine Plattform oder einen Marktplatz zur Verfügung. Sobald ein Bieter den Zuschlag für ein Geschäft erhält, bietet sich für eine sichere Zahlung der Transaktionsbroker (Transaction Broker) an, eine weitere Subform des Brokerage models. Der Transaction Broker stellt einen vertrauenswürdigen Zahlungsmechanismus bzw. -kanal zur Verfügung, der sowohl für den Verkäufer als auch für den Bieter eine sichere Zahlungsabwicklung ermöglicht.

9.4.2 Advertising Model

Definition: Das Advertising Model repräsentiert den werbefinanzierten Ansatz des klassischen Mediengeschäftes und bietet kostenlose Inhalte und Produkte im Austausch gegen die Bereitschaft, Werbung zu konsumieren.

Erlösform: indirekt, via Unternehmen, Werbung

Das Advertising Model stellt ein sehr bedeutendes Geschäftsmodell der Internetökonomie dar. Das Modell repräsentiert den werbefinanzierten Ansatz des klassischen Mediengeschäftes und bietet kostenlose Inhalte und Produkte im Austausch gegen die Bereitschaft, Werbung zu konsumieren. Beim Advertising reichen die Möglichkeiten der Werbefinanzierung von einem "'subventionierten"' Verkauf, zum Beispiel bei Online-Zeitungsartikeln, bis zu einer für den Konsumenten scheinbar kostenlosen Weitergabe von Inhalten, wie bei der Weitergabe von Informationen und digitalen Produkten. "'Die Generierung von Reichweiten ist eines der wichtigsten Ziele für Medien und Kommunikationsunternehmen, um im Komplementärmarkt der Werbung möglichst hohe Umsätze erzielen zu können."' [14] Beispiele für erfolgreiches Advertising bieten Web-Portale wie z.B. bluewin.ch, die dem Kunden ein sehr breites Spektrum verschiedenster Inhalte anbieten, wodurch ein grosses Zielpublikum (grosse Reichweite, bzw. Quantität der Rezipienten) angesprochen werden kann. Dies wiederum ermöglicht die Refinanzierung der angebotenen Inhalte zu einem grossen Teil durch Werbung.

9.4.3 Infomediary Model

Definition: Das Infomediary Model bietet ähnlich wie die werbefinanzierten Modelle kostenlose Inhalte und Produkte, im Gegensatz dazu aber nur im Austausch gegen persönliche Daten der Kunden.

Erlösform: indirekt, via Unternehmen, Datamining

Infomediäre (information intermediary) zeichnen Benutzerdaten auf, die durch Nutzertransaktionen jedweder Art anfallen, wie z.B. bei Gewinnspielen, Fragebögen, etc.. Dieses Sammeln von Daten wird auch als Data Warehousing bezeichnet. Die Möglichkeit des Trackings bietet dem Internet Vorteile gegenüber anderen Medien. Sog. Log files zeichnen verschiedene Nutzerdaten auf, wie z.B. Nutzung der Hauptseite und der Einzelseiten, also die Visits oder Pageviews pro Zeiteinheit, die Herkunfts-Domain des Nutzers, das installierte Betriebssystem und der verwendete Browser, werbemittelbezogene Daten (Click-Through), sowie Links die Traffic für die jeweilige Webseite generieren. Ausserdem sind die sog. "'ZAG-Daten"' sehr beliebt. Die Abkürzung ZAG steht für Zip Code, Age und Gender, also für die Postleitzahl, für das Alter und das Geschlecht.

In einem zweiten und weitaus aufwändigeren Schritt werden die gesammelten Daten aufbereitet und analysiert, um Konsumgewohnheiten und Daten der Konsumenten herauszuarbeiten. Dieser zweite Schritt wird auch Datamining genannt und bezeichnet im Gegensatz zum Data Warehousing nur die Auswertung und Aufbereitung der Daten. Die Refinanzierung erfolgt in einem dritten Schritt durch die Weitergabe oder die Bereitstellung der Daten an dritte Unternehmen, die mit den Informationen neue Kunden ansprechen oder ihr Leistungsangebot optimieren möchten. Data Warehousing und Datamining sollen also eine zielgruppengerechte Ansprache - One-to-One genannt - der Nachfrager durch die Anbieter ermöglichen. Das Infomediary Model ist damit ein indirektes Modell, mit grossem Potential in der Internet-Ökonomie. Die Bereitschaft der Anbieter, sich für Daten über Konsumenten und deren Kaufgewohnheiten an Infomediäre zu wenden, liegt in der Minimierung der Streuverluste der Anbieter, wodurch sie die begrenzten Mittel für Werbung gezielter einsetzen können.(siehe dazu [14], S. 168)

9.4.4 Merchant Model

Definition: Das Merchant Model stellt ein klassisches Handelsgeschäft in elektronischer Form dar, das durch Sortimentsgestaltung im Internet Mehrwert schafft.

Erlösform: direkt, nutzungsunabhängig, einmalig

Unter dem Begriff Merchant Model versteht man generell das Handeln von Gütern und Dienstleistungen der Gross- und Einzelhändler in elektronischen Netzwerken. In den einzelnen Ausprägungen wird weiter unterschieden, welche Art von Gütern veräussert werden und über welche Kanäle diese Güter den Kunden angeboten und ausgeliefert werden. So bietet ein Virtual Merchant seine Ware (Soft-/Hard-Goods) (siehe dazu [5], S. 394) ausschliesslich über das Internet an. Demgemäss kann die Ware auch nur über das Internet bestellt werden. Ein Bit Vendor verkauft ausschliesslich Produkte in digitaler Form, also Soft-Goods, die wiederum nur über das Internet bestellt und ausgeliefert werden können. Der Catalog Merchant hingegen bietet seine Produkte über einen webbasierten Katalog an, die aber nicht nur über das Internet bestellt werden können. Click and Mortar ist

ein weiteres Submodel des Merchant Models, bei dem ein physisch existierender Einzelhändler seine Produkte als eine zusätzliche Dienstleistung neben dem normalen Verkauf in physisch existierenden Warenhäusern über das Internet anbietet.

9.4.5 Manufacturer (Direct) Model

Definition: Das Manufacturer Model repräsentiert den Direktvertrieb eines Herstellers an den Endkunden und basiert auf der Ausschaltung der Zwischenhändler.

Erlösform: direkt, nutzungsunabhängig, einmalig

Das Manufacturer Model ermöglicht es den Herstellern, ihre Waren und Dienstleistungen direkt an den Kunden weiterzugeben, ohne auf Zwischenhändler angewiesen zu sein. Dadurch können Kosten eingespart, und das Verhältnis zwischen dem Kunden und dem Anbieter intensiviert werden. Weil der Hersteller (Manufacturer) sein Produkt auf seine Kunden abstimmen kann (Mass Customization), können die Kundeninteressen und -bedürfnisse besser berücksichtigt werden. Im Vergleich zum herkömmlichen Merchant werden die Produkte beim Value Chain Integrator erst nach einer erfolgten Bestellung hergestellt. Eine eigentliche Lagerhaltung, sowie Ladenlokale im herkömmlichen Sinne bestehen nicht. Der Manufacturer vertreibt seine Ware fast ausschliesslich über das Internet.

9.4.6 Affiliate Model

Definition: In dem Affiliate Model wird das Affiliate Programm als Geschäftsmodell institutionalisiert, das auf der Vermittlung von Kundenkontakten basiert.

Erlösform: indirekt, via Unternehmen, Kommission

Eine weitere Erlösform ist die Kommission, die vor allem bei Partnerprogrammen zum Einsatz kommt. E-Commerce-Anbieter, bislang vorwiegend aus dem Buch oder Musikbereich, suchen sich reichweitenstarke und zielgruppenaffine Online-Anbieter, die sog. "'Affiliates"'. Auf deren Seiten im Internet werden Werbungen in Form von Bannern oder Buttons geschaltet, die den User auf die Site des E-Commerce-Anbieters weiterleiten. Tätigt der weitergeleitete User einen Kauf, muss der E-Commerce-Anbieter einen vereinbarten Prozentsatz, also eine Kommission des Kaufumsatzes an den Affiliate rückvergüten.

Dadurch, dass das absolute Volumen der Werbeausgaben im Internet im Vergleich mit anderen Medien immer noch gering ist, bieten Affiliate-Programme eine mögliche Erlösquelle für Online-Medien aus dem Special-Interest-Bereich. Die geringe Reichweite und die daraus resultierenden niedrigen Werbeeinnahmen können teilweise durch das Angebot einer spezifischen Zielgruppe und daraus resultierenden Kommissionen kompensiert

werden. Für den E-Commerce-Anbieter spielt vor allem der kontextbezogene Inhalt des Online-Publishers zu seinen angebotenen Gütern oder Dienstleistungen eine wichtige Rolle. Vertriebspartnerschaften können gerade im Internet zu einer deutlichen Steigerung der Einnahmen von kommerziellen Websites führen.

9.4.7 Community Model

Definition: Das Community Model nutzt die Netzwerkeffekte und Loyalität von Interessensgemeinschaften und refinanziert sich durch das ökonomische Potential der Bündelung von Kundenbedürfnissen.

Erlösform: indirekt, via Kunden/Unternehmen, "'Sponsoring"'

Die Realisierbarkeit des Community Models steht und fällt mit der Loyalität der Internetnutzer. Umsatz wird durch den Verkauf von ergänzenden Produkten und Dienstleistungen oder durch freiwillige Beiträge erzielt.

9.4.8 Subscription Model

Definition: Das Subscription Model basiert auf einem Einheitspreis für beliebig viele Einheiten eines Gutes (Pauschalpreis) und bietet somit eine nutzungsunabhängige Bezahlung von Dienstleistungen oder Informationen.

Erlösform: direkt, nutzungsunabhängig, regelmässig wiederkehrend, Abonnement

Das Subscription Model bezeichnet all jene Geschäftsformen im Internet, bei welchen für den Abruf von Inhalten jedweder Form periodisch Abonnementsgebühren verrechnet werden. Die Erlösform des Abonnements eignet sich für zwei Zielgruppen. Erstens für Spezialanbieter von Fachinformationen, deren Nutzer vor allem aus dem Geschäftsbereich stammen. Zweitens für Special-Interest-Inhalte, die auf ein hohes Interesse der Nutzer stossen. Beispiele dafür sind Abonnementgebühren für Online-Gaming, Erotic-Channels, oder Sport-News(siehe dazu [14], S. 171)

Während Abonnementgebühren in den traditionellen Medien wie Printmedien oder Fernsehen zu den wichtigsten Erlösströmen gehören, haben sich Abonnementgebühren im Internet nur teilweise durchsetzen können. Gemäss dem Motto "'follow the free"' zeigen sich User nur zögerlich bereit, für Inhalte aus dem Internet Gebühren zu bezahlen. Dies mag durch die Entstehung des Internets bedingt sein. Generell kann man aber feststellen, dass die Zahlungsbereitschaft der Internetbenutzer stetig zunimmt, auch für allgemeine Inhalte.

9.4.9 Utility Model

Definition: Das Utility Model ist das Gegenstück zum Subscription Model, denn es basiert auf der nutzungsabhängigen Bezahlung von Dienstleistungen oder Informationen.

Erlösform: direkt, nutzungsabhängig, Einzeltransaktion nach Leistungsmenge und/oder Leistungsdauer.

Das Utility Model bezeichnet all jene Geschäftsformen im Internet, bei welchen für den Abruf von Inhalten jedweder Form Gebühren nach Leistungsmenge und/oder -dauer verrechnet werden. Es handelt sich um ein bekanntes Abrechnungssystem, das sich bereits bei alltäglichen Dienstleistungen wie Elektrizität, Wasser oder Ferngesprächen bewährt hat. Viele Internet Service Provider berechnen ihre Dienste nach Leistungsdauer und/oder Leistungsmenge.

Das Utility Model hat sich schon vor der e-Commerce-Aera bewährt, was sich positiv auf die Kundenakzeptanz des Modells auswirkt. Für Zahlungen im Micropayment-Bereich ist das Vorhandensein von geeigneten elektronischen Bezahlsystemen der entscheidende kritische Erfolgsfaktor für das Utility model, um nachhaltig und auf breiter Front Umsätze generieren zu können.

9.5 Fallstudien

9.5.1 iTunes

Geschäftsmodell Das Geschäftsmodell von iTunes (siehe Abbildung 9.7) basiert auf dem Merchant Model von Rappa (vgl. Kapitel 1.4.4). iTunes bietet eine Auswahl von über einer Mio. Musiktiteln, woran 4 der grossen Platten-Labels (EMI, Sony/BMG, Universal und Warner Bros) beteiligt sind. Der Preis in den USA beträgt $0.99 pro Titel bzw. $9.99 für ein ganzes Album. Mit dem Kauf erwirbt man sich die Rechte, die Titel für den persönlichen Gebrauch auf CD zu brennen, auf bis zu drei Computern und einer unlimitierten Anzahl von tragbaren iPods - dem MP3-Player von Apple - zu hören. Die Abrechnung erfolgt somit gemäss der Anzahl bezogener Musiktitel. Mit der iTunes Software können mittlerweile auch Radiostationen und Hörbücher gehört und sogar Film-Trailer betrachtet werden. (vgl. [16] und [18])

Alternative Modelle Alternativ denkbar wäre eine Variante mit einer monatlichen Gebühr nach dem Subscription Model oder ein Utility Model, das nur eine beschränkte Anzahl von Nutzungen erlaubt.

Figure 9.7: www.itunes.com (Stand am 12. Januar 2005)

Firmengeschichte und Entwicklung Die Musikplattform iTunes.com wurde im Jahr 2003 von Apple in Ergänzung zur vorhandenen iTunes Software und dem MP3-Player iPod lanciert. In der ersten Betriebswoche wurden mehr als 1 Million Musiktitel zu $0.99 heruntergeladen. (vgl. [15], S. 100)

Apples iTunes Software (siehe Abbildung 9.8) enthält einen Musik-Player, CD Ripping und Brenn-Tools, ein Interface zum iPod, gratis Internet-Radiostationen und eine limitierte Streaming-Audio-Funktion namens Rendezvous.

Im Oktober 2003 wurde eine entsprechende Software für Windows angekündigt, sowie weitere Partnerschaften mit AOL und Pepsi Cola eingegangen.

Zuerst war das Musikportal nur in den USA verfügbar, weitere internationale Stores werden nun aber fortlaufend eröffnet, beispielsweise in Grossbritanien, Irland, Frankreich, Deutschland, Kanada, Österreich, Belgien, Finnland, Griechenland, Italien, Luxemburg, Niederlande, Portugal und Spanien.

Marktposition und -anteil Gemäss eigenen Angaben von Apple [17] wurde im Dezember 2004 die Marke von 200 Millionen verkauften Musiktiteln überschritten. Dies macht iTunes zum wohl erfolgreichsten Online-Musikanbieter überhaupt.

Zukunftsaussichten Der Erfolg von iTunes spricht für sich (Marktanteil von 70%). Insbesonders können Netzwerkeffekte mit dem iPod und anderen Produkten von Apple ausgenutzt werden. Es stellt sich die Frage, ob damit dem Trend von MP3-Raubkopien via P2P-Netzwerke (wie z.B. Kazaa) entgegengewirkt werden kann.

Figure 9.8: Apples iTunes Store

9.5.2 Yahoo!

Geschäftsmodell Yahoo! (siehe Abbildung 9.9) ist eines der ersten und wohl bekanntesten klassischen Internetportale und ist in das Advertising Model von Rappa einzuordnen.

Neben dem redaktionell aufbereiteten Verzeichnis von Internetseiten und der damit verbundenen Suchfunktion stehen weitere Sevices zur Verfügung. Die Dienste können mittels MyYahoo! personalisiert werden und zu einem essentiellen Teil im Leben der Nutzer werden.

Kommunikationsdienstleistungen sind z.B. Yahoo! Mail, Yahoo! Messenger, Yahoo! Calendar, Yahoo! Chat, Yahoo! Greetings, Yahoo! Clubs oder Yahoo! Photos. Desweiteren werden verschiedene interessenspezifische Portale angeboten, u.a. Yahoo! Shopping, Yahoo! Auctions, Yahoo! Finance und Yahoo! Travel. Zudem sind weitere Inhalt- und Medienprogramme zu verschiedensten Interessen verfügbar (wie z.B. Yahoo! Sports, Yahoo! Music, Yahoo! Movies, Yahoo! News und Yahoo! Games). [19] Das Werbeangebot für Anbieter umfasst klassische Banneranzeigen bis hin zu neuen Modellen, wie z.B. gesponserten Inhalten.

Alternative Modelle Alternativ denkbar wären z.B. eine Suchbox anhand eines Affiliate-Models, die auf einzelnen Internetseiten plaziert werden kann. Der Betreiber der entsprechenden Internetseite könnte dann z.B. pro überwiesenem Nutzer eine Art "'Kopfgeld"' erhalten. Eine andere Variante wäre eine Abrechnung gemäss Anzahl Suchanfragen nach dem Utility-Model. Obwohl der Erfolg dieses Modells fraglich ist.

Figure 9.9: www.yahoo.com (Stand am 12. Januar 2005)

Firmengeschichte und Entwicklung Yahoo! wurde im Jahr 1994 von den beiden Standford Doktoranden David Filo und Jerry Yang gegründet und starte ursprünglich als "'Jerryś Guide to the World Wide Web"'. Der Name Yahoo! ist ein Akronym für "'Yet Another Hierarchical Officious Oracle"'. Im Herbst 1994 konnte der erste Tag mit 1 Mio. Hits gefeiert werden, dies entsprach rund 100'000 täglichen Besuchern. Offiziell wurde Yahoo! Inc dann im April 1995 mit einem Gründungskapital von 2 Mio. Dollar gegründet. Das Managementteam wurde fortlaufend erweitert, so z.B. mit Tim Koogle, einem Motorola-Veteranen, oder Jeffrey Mallett, einem Gründer von Novells WordPerfect Kundenabteilung. Im April 1996 ging Yahoo! an die Börse, damals mit einem Total von 49 Angestellten. Heute erreicht Yahoo! rund 232 Mio. Nutzer monatlich und besitzt weltweit rund 25 Niederlassungen. [20]

Marktposition und -anteil Yahoo! ist - gemäss eigenen Angaben [19] eine der meisterkanntesten Marken weltweit, die 237 Mio. Nutzer in 25 Ländern und 13 Sprachen erreicht. Yahoo! stellt die Nummer 1 aller Internetmarken dar und erreicht die meisten Nutzer. Mittels Yahoo! Company - einer Portal-Lösung für Firmen konnte auch im wachsenden Business-Bereich Fuss gefasst werden.

Zukunftsaussichten Es wird sich zeigen, ob sich Yahoo! im Suchmaschinenmarkt behaupten werden kann. Neue Suchmaschinen wie Google & Co stellen hier eine grosse Konkurrenz dar.

Im Portalgeschäft wird sich - unserer Meinung nach - Yahoo! behaupten können. Vor allem können aufgrund der hohen Zugriffszahlen Informationen zu Trends ermittelt wer-

den. Zum Beispiel können anhand der Zugriffszahlen die einzelnen Angebote und Unterabschnitte seitens Yahoo optimiert werden. Andererseits stellen gerade hochfrequentierte Internetseiten erst einen attraktiven Markt für Werbende dar.

9.5.3 Amazon.com

Figure 9.10: www.amazon.com (Stand am 12. Januar 2005)

Geschäftsmodell Amazon.com (siehe Abbildung 9.10) ist gemäss Rappa in verschiedenen Gebieten tätig: als virtueller Marktplatz mit automatisierten Transaktions- und CRM-Services (Customer Relationship Management) agiert Amazon.com im Sinne des Brokerage Models von Rappa. Spezialisierte Kaufportale für verschiedene Interessen in Zusammenarbeit mit Drittanbietern können mittels dem Merchant Model klassifiziert werden. Zudem verfügt Amazon.com über sein berühmtes Affiliate-Programm (Affiliate Model nach Rappa) womit die Möglichkeit besteht von externen Seiten auf bestimmte Produkte hinzuweisen. Die Abrechnung erfolgt per Umsatzbeteiligung.

Firmengeschichte und Entwicklung Amazon.com hat seinen Standort in Seattle und wurde von Jeff Bezos gegründet. Die Webseite ging 1995 online, der Börsengang erfolgte im Mai 1997. Als klassischer Vertreter der eBusiness-Firmen startete auch Amazon.com zuerst in einer Garage und wurde fortlaufend vergrössert. Die Vision von Amazon.com kann anhand des folgenden Zitates verdeutlicht werden:

> "'Our goal is to be Earth's most customer-centric company. I will leave it to others to say if we've achieved that. But why? The answer is three things:

> The first is that customer-centric means figuring out what your customers want by asking them, then figuring out how to give it to them, and then giving it to them. That's the traditional meaning of customer-centric, and we're focused on it. The second is innovating on behalf of customers, figuring out what they don't know they want and giving it to them. The third meaning, unique to the Internet, is the idea of personalization: Redecorating the store for each and every individual customer. If we have 10.7 million customers, as we did at the end of the last quarter, then we should have 10.7 million stores."' [21]

Marktposition und -anteil Amazon.com besitzt - ausser der internationalen Verkaufsplattform amazon.com - verschiedene lokale Online-Stores wie in Grossbritanien (Amazon.co.uk), Deutschland/Österreich/Schweiz (Amazon.de), Frankreich (Amazon.fr), Japan (Amazon.co.jp) oder Kanada (Amazon.ca).

Zukunftsaussichten Amazon.com ist heute die grösste, wirklich erfolgreiche reine Internet-Buchhandlung weltweit. Durch die Grösse können Effekte wie Economies of Scale ausgenützt werden, was entscheidende Vorteile gegenüber der Konkurrenz bietet. Gerade in diesem Geschäftsfeld mit sehr niedrigen Margen sind diese notwendig, um überhaupt Gewinn zu erwirtschaften.

9.5.4 Reuters

Figure 9.11: www.reuters.com (Stand am 12. Januar 2005)

Geschäftsmodell Reuters (siehe Abbildung 9.11) bietet weltweit speziell aufbereitete Information, die auf Finanzdienstleistungen, Medien und auch auf Firmen zugeschnitten ist. Schnelligkeit, Richtigkeit sowie Unbefangenheit sind dabei Eigenschaften durch welche Reuters bekannt geworden ist. Reuters ist somit ein Vertreter des Subscription Model von Rappa und kann zudem auch als ein Content Provider im engeren Sinn verstanden werden. Je nach Abrechnung (aufgrund der Anzahl Inhalte) könnte auch das Utility Model angewendet werden.[25]

Firmengeschichte und Entwicklung 1851 eröffnete Paul Julius Reuter in London ein Büro, das Börseninformationen zwischen London und Paris übermittelte. Dieser Service wurde später auf die britische Presse und weitere europäische Länder ausgeweitet. So war Reuters z.B. 1865 die erste Informationsstelle, die über die Ermordung des amerikanischen Präsidenten Lincoln berichtete. Mit der technologischen Weiterentwicklung von Telegraphen und Unterwasserkabeln wurde das Netzwerk von Reuters fortlaufend ausgeweitet. In einigen Bereichen wie der Funkübertragung war Reuters entscheidend beteiligt.

In der zweiten Hälfte des 20. Jahrhunderts erfolgte eine Modernisierung der Infrastruktur. 1984 erfolgte der Börsengang. Reuturs wuchs fortan kontinuierlich weiter, die Angebotspalette wurde um zahlreiche Businessangebote und Dienstleistungen im Medien-, Finanz- und Wirtschaftsbereich ausgebaut. Im Zuge des Internetbooms erfolgten zahlreiche Investitionen in Start-up Firmen und das Angebot wurde auf Internetdienstleistungen ausgeweitet.[26]

Marktposition und -anteil Reuters ist heute der Inbegriff von aktueller, akurater Information. Der Marktanteil betrug im Jahr 2003 rund 37% [27]. Reuters ist durch das Anbieten einer grundlegenden Sache wie Information unabhängig von Hypes und Börsencrashes. Die Inhalte werden lediglich über neue Vertriebskanäle wie das Internet verbreitet, an der eigentlichen Dienstleistung ändert sich nichts.

Zukunftsaussichten Reuters wird sich gegenüber anderen Informationsanbietern wie z.B. Bloomberg behaupten müssen. Aktuelle, akurate Information wird nach wie vor gefragt sein.

9.6 Schlussfolgerungen

9.6.1 Fazit

Die Entstehung des Internets als eine neue Infrastruktur hat zur Veränderung der bisherigen Vertriebs- und Handelsstrukturen geführt und hat zugleich neue ökonomische Marktmodelle mit neuen Spielregeln erzeugt, die insbesondere für Unternehmen aus dem Medien- und Kommunikationssektor gelten. Traditionelle und neu entstandene Inhalte jeglicher

Art werden im World Wide Web mit unterschiedlichen Geschäftsmodellen über verschiedene Kanäle angeboten und veräussert, so dass sich eine Übersicht und eine Analyse über das Content Provider Business geradezu aufdrängt. Die vorliegende Arbeit hatte deshalb das Ziel, ein Klassifikationsschema des Content Provider Business zu bieten, d.h. eine mögliche Gliederung der zurzeit verwendeten Geschäftsmodelle internetbasierter Unternehmen. Ausserdem sollten die Modelle in einem weiteren Schritt anhand von Fallstudien diskutiert und analysiert werden.

Dafür musste in einem ersten Teil der Arbeit der schwer zu fassende Begriff "'Content Provider Business"' aufgegriffen und definiert werden. In der vorhandenen Literatur lassen sich zwei unterschiedliche Ansätze für eine mögliche Definition erkennen. Zum einen die engere Auffassung, unter welcher ausschliesslich die Produktion der Inhalte verstanden wird und zum zweiten die weiter gefasste Definition, die neben der Erschaffung der Inhalte auch die Aufbereitung, Veredelung und Distribution miteinbezieht. Für die Arbeit sind wir von der zweiten Definition ausgegangen und haben sie weiter angepasst. Wir sind der Meinung, dass auch weiterführende Informationen über angebotene Produkte im weiteren Sinne als Inhalt zu verstehen sind. Darunter fallen Beschreibungen von angebotenen Gütern durch den Anbieter und durch Kundenrezensionen und zusätzliche Informationen, die beim Kunden einen Mehrwert schaffen können. Diesem Verständnis von Content Provider Business folgend, konnte ein grösseres Betrachtungsfeld untersucht werden, das mit wenigen Ausnahmen das ganze Spektrum des e-Commerce berührt. Dafür musste aber auch ein Klassifikationsschema gefunden werden, welches dieses Spektrum abdecken kann.

Eine weitere Schwierigkeit liegt im Begriff "'Geschäftsmodell"', der in der Forschungsliteratur sehr unterschiedlich behandelt wird. Wir verstehen unter dem Begriff ein zeitpunktbezogenes, aggregiertes Modell, das die Akteure eines Geschäftsnetzwerkes, ihre Rollen und Austauschbeziehungen sowie ihrer Nutzenpotentiale aus Sicht eines fokussierten Akteurs beschreibt. [4]

Für die Klassifikation der Geschäftsmodelle haben wir auf die Gliederung von Michael Rappa zurück gegriffen, die von neun Grundmodellen ausgeht und sich in weitere vierzig Submodelle aufteilen lässt. Das Klassifikationsschema orientiert sich weitgehend an der Praxis, und die einzelnen Modelle werden durch textuelle Beschreibungen erläutert, wie in Kapitel 1.4 beschrieben.

Für eine präzisere Beschreibung schien es ausserdem notwendig, weitere Kriterien für die Gliederung der Modelle zu berücksichtigen. So wurde der Ansatz von Rao übernommen, der die beteiligten Akteuere eines Geschäftsnetzwerkes in Business- und Customer-Bereich unterteilt. Eine präzise Beschreibung der Erlösform für jedes der neun Grundmodelle stellt ein zusätzliches wichtiges Kriterium dar. Dafür haben wir die Erlösform von Zerdick übernommen und diese leicht für den Content-Bereich angepasst.

Die Einteilung der Fallstudien in die ausgesuchte Klassifikation erwies sich dennoch als ein schwieriges Unterfangen. Für die meisten Inhaltsanbieter wie iTunes, Yahoo etc. lassen sich mehrere Grundmodelle ausmachen, je nachdem welchen Geschäftsbereich der Fallstudien gerade fokussiert wird. Bei den vierzig Submodellen verhält es sich ähnlich: Bei einigen Modellen ist eine klare Abgrenzung nicht möglich. Viele der Modelle ergeben gerade erst in Kombination mit anderen Modellen einen Sinn, so kann z.B. das Affiliate

Model nur dann angewendet werden, wenn dem Kunden tatsächlich auch noch eine Ware oder eine Dienstleistung angeboten wird.

Eine fundierte Analyse war im Rahmen dieser Arbeit nicht zu leisten. Dafür ist der gewählte Bereich des Content Provider Business zu weit gefasst.

9.6.2 Ausblick

Nach einer ersten Konsolidierung im Internet-Markt ist die grosse Euphorie der New Economy einer kritischeren Haltung gewichen. Modelle, die nur auf Werbeeinnahmen basieren, sind auf lange Frist nicht tragbar, vor allem, wenn dem Anbieter bei der Produktion und der Bereitstellung der Inhalte hohe Kosten anfallen. Selbst Geschäftsmodelle mit grossem Wachstumspotential helfen nichts, wenn auf lange Sicht gesehen, keine schwarzen Zahlen geschrieben werden können. Solange die Internet-Nutzer nicht bereit sind, für allgemeine Inhalte zu bezahlen, und solange sich das absolute Werbevolumen auf einem tiefen zweistelligen Bereich befindet, muss die Refinanzierung der Geschäftstätigkeit über weitere Modelle erfolgen.

Das Internet ermöglicht auch in Zukunft die Bildung neuer Interessensgruppen und deren ökonomisch attraktive Versorgung mit medialen Inhalten. Vor allem im Bereich der Mesomedien, d. h. Interessensgruppen von bis zu 10000 Personen liegt grosses Potential, das mit den traditionellen Medien nicht abgedeckt werden kann. Neue Geschäftsmodelle internetbasierter Unternehmen bieten Möglichkeiten an, diese rentabel abzudecken. [14]

Ein nicht zu unterschätzender Bereich in der Internet Ökonomie ist die Spiele-Industrie, die durch geschickte Konzepte immer mehr zahlungsbereite Kunden in die Welt der Online-Spiele zu locken vermag. Gerade das Subscription Model dürfte dabei eine wichtige Rolle spielen, und das obwohl Abonnementgebühren im Internet lange als unmöglich zu realisieren galten. Aufgrund mangelnden Erfolges haben in einer Frühphase des Internets viele Anbieter ihre Abonnementgebühren aufgegeben. Während Abonnementgebühren in den traditionellen Medien wie Printmedien oder Fernsehen zu den wichtigsten Erlösströmen gehören, haben sich Abonnementgebühren im Internet bis jetzt nur teilweise durchsetzen können. Generell lässt sich in den letzten Jahren aber feststellen, dass die Zahlungsbereitschaft der Internetnutzer stetig zunimmt und dass sich durchaus auch Modelle durchsetzen können, die auf Abonnementen basieren.

Als ein weiteres Mittel zur Erlösgenerierung ist das Datamining zu sehen, das eine zielgruppengerechte Ansprache der Nachfrager durch die Anbieter ermöglichen soll. Obwohl sich die Hoffnungen in das One-to-One-Marketing bis jetzt noch nicht erfüllt haben, könnte darin die Zukunft der Internet-Ökonomie liegen. Datamining wird immer stärker dazu eingesetzt, Kunden mit speziell auf sie ausgerichtete Inhalten an die Anbieter zu binden. Für das sog. "'Tracking"', also die Nachverfolgung der Kunden ist das World Wide Web ein geeignetes Mittel, da Kundendaten leicht zugänglich und ohne grossen Aufwand speicherbar sind. Probleme bietet allerdings die Aufbereitung und Handhabung der Daten. Da sich viele Online-Anbieter immer noch in der Aufbauphase befinden, können nur ungenügend Ressourcen für Infomediäre und Datamining zur Verfügung gestellt werden, und die aufgezeichneten Daten werden grösstenteils ungenutzt wieder gelöscht.

Bibliography

[1] Allan Afuah, Christopher L. Tucci, Internet Business Models and Strategies, 2001

[2] Bambury, Paul: A Taxonomy of Internet Commerce. Online im Internet: http://www.firstmonday.dk/issues/issue3_10/bambury/index.html (Stand 12.01.2005)

[3] Bayers, Charles: The Promise of One to One (A Love Story). In: Wired, Nr. 5/1998 S. 130-134

[4] Breuer, Steffen Beschreibung von Gesch¨aftsmodellen internetbasierter Unternehmen. Konzeption - Umsetzung - Anwendung (Diss. HSG St. Gallen, Nr. 2923), Bamberg 2004.

[5] Merz M., E-Commerce und E-Business: Marktmodelle, Anwendungen und Technologien, Heidelberg, 2002.

[6] Porter M. und Millar V.E.: How Information Gives You Competitive Advantage, in: Harvard Business Review, Nr. 4, 1985. S. 149-160

[7] Bharat Rao, Emerging Business Models in Online Commerce, 1999. Online im Internet: http://www.ite.poly.edu/people/brao/RT99.pdf (Stand 12.01.2005)

[8] Slywotzky, A.: Value migration: how to think several moves ahead of the competition, Boston, 1996

[9] Wired.com: Webvan. Online im Internet: http://www.wired.com/news/business/0,1367,45098,00.html (Stand 12.01.2005)

[10] Timmers, P: Business Models for Electronic Markets, 1998. In: EM - Electronic Markets, Vol. 8, No.2, 1998.

[11] Weill, Peter; Vitale, Michael R.: Place to Space - Migrating to eBusiness Models, Boston, 2001.

[12] Wimmer, Helmut: Zur Konvergenz von Technologie und Wissen, Diplomarbeit Universität Wien, 1997

[13] Wirtz, B.W.; Kleineicken, A.: Geschäftsmodelltypologien im Internet. In: WiSt, Heft 11, 2000, S. 628-635

[14] Zerdick A. et al.: Die Internet-Ökonomie: Strategien für die digitale Wirtschaft, Berlin, 2003.

[15] Kahney, Leander: Wired.com, Apple Launches Paid Music Service. Online im Internet: http://www.wired.com/news/business/0,1367,58656-2,00.html (Stand am 12.01.2005).

[16] Berkman Center for Internet & Security at Harvard Law School: iTunes: How Copyright, Contract, and Technology Shape the Business of Digital Media. Online im Internet: http://cyber.law.harvard.edu/media/uploads/53/GreenPaperiTunes03.04.pdf (Stand 12.01.2005)

[17] Apple: iTunes Music Store Downloads Top 200 Million Songs. Online im Internet: http://www.apple.com/pr/library/2004/dec/16itunes.html (Stand 12.01.2005)

[18] Apple: The #1 music download store. Online im Internet: http://www.apple.com/itunes/store/ (Stand 12.01.2005)

[19] Yahoo! Inc: Company Overview. Online im Internet: http://docs.yahoo.com/info/misc/overview.html (Stand 12.01.2005)

[20] Yahoo! Inc: The History of Yahoo! - How It All Started... Online im Internet http://docs.yahoo.com/info/misc/history.html (Stand 12.01.2005)

[21] Krishnamurthy, Sandeep: Amazon.com - A business history, 2002. Online im Internet: http://www.swlearning.com/marketing/krishnamurthy/first_edition/case_updates/amazon_final.pdf (Stand 12.01.2005)

[22] Rappa, Michael: Business models on the web, Managing the digital enterprise, 12.01.2004. Online im Internet: http://digitalenterprise.org/models/models.html (Stand am 01.12.2004).

[23] Hedman J., Kalling T., The business model concept: theoretical underpinnings and empirical illustrations, in: European Journal of Information Systems, Nr. 1, Jg. 12, 2003 S. 49-59

[24] Kahney, Leander: Wired.com, Apple Launches Paid Music Service. Online im Internet: http://www.wired.com/news/business/0,1367,58656-2,00.html (Stand am 12.01.2005).

[25] Reuters: Company Overview. Online im Internet: http://about.reuters.com/aboutus/overview/ (Stand am 12.01.2005).

[26] Reuters: Reuters History. Online im Internet: http://about.reuters.com/aboutus/history/ (Stand am 12.01.2005).

[27] Finextra.com: Reuters overall market share slips, high-end gains on Bloomberg. Online im Internet: http://www.finextra.com/fullstory.asp?id=11557 (Stand am 12.01.2005).

Chapter 10

Verrechnungsmethoden für Inhalte aus dem Internet

Franc Uffer, Michael Biedermann, Roger Loosli

Mit der weltweiten Verbreitung des Internets stellt sich die Frage nach einer geeigneten Verrechnungsmethode für Internetinhalte immer mehr. Seien es Artikel von Onlinemagazinen oder Softwarebestellungen, die herkömmlichen Abrechnungsmethoden wie Rechnung oder Banküberweisung können mit der Geschwindigkeit des Internets nicht mithalten. Die heute für Internetzahlungen breit verwendete Kreditkarte stösst aufgrund Sicherheitsbedenken und teils hoher Kommissionen vor allem im Micropayment Bereich, also Zahlungen bis etwa 10 Schweizer Franken an ihre Verwendungsgrenzen. Die Nachfrage nach einer einfachen, anonymen und möglichst universalen Verrechnungsmethode ist also sowohl auf Seiten der Händler wie auch auf der potenziellen Kunden gross. In dieser Arbeit werden verschiedene Methoden vorgestellt und miteinander verglichen. Dabei werden auch Gründe für das Scheitern von neuartigen Zahlungssystemen anhand ausgewählter Beispiele aufgezeigt. Anschliessend werden die erfolgskritischen Anforderungen an eine Zahlungsmethode erarbeitet und anhand derer die Marktchancen existierender Systeme diskutiert. In einer abschliessenden Diskussion wird das Erarbeitete kurz präsentiert und es wird versucht, realistische Schlüsse für die Zukunft zu ziehen.

Inhaltsangabe

10.1 Übersicht über aktuelle Verrechnungsmethoden

In der vorliegenden Arbeit werden in einem ersten Schritt verschiedene aktuelle Systeme vorgestellt und miteinander verglichen. Anschliessend wird untersucht, weshalb sich viele zum Teil technisch ausgereifte Systeme nicht oder nur lokal durchzusetzen vermochten. Abschliessend werden die erfolgskritischen Anforderungen an Bezahlsysteme erarbeitet und anhand dieser die Marktchancen aufkommender Methoden abgeschätzt. In einer zusammenfassenden Beurteilung wird danach versucht, aus dem Gezeigten Schöüsse zu ziehen und einen Ausblick auf die nächsten Jahre zu tätigen.

10.1.1 Einleitung

Seit der Mensch begonnen hat sich zu spezialisieren und nicht mehr alles selbst herstellte, wurde ein Bezahlsystem für den Handel notwendig. Anfangs bestand dieses System im simplen Tauschhandel. Mit der Erfindung des Geldes wurde jedoch ein Bezahlsystem erreicht, welches im Grossen und Ganzen auch heute noch besteht. Durch die immer stärker werdende elektronische Vernetzung im 21. Jahrhundert besteht eine grosse Nachfrage nach einem Bezahlsystem, welches mit den speziellen Eigenschaften der virtuellen Welt kompatibel ist. Vor allem bei kleineren Beträgen sollen sich die Vertragsparteien nicht mehr um Kleinigkeiten wie Währungsumrechnnug oder Standort der Parteien kümmern müssen.

10.1.2 Übersicht über aktuelle Verrechnungsmethoden

Um einen ersten Überblick zu erhalten werden im Folgenden aktuelle Verrechnungsmethoden kurz vorgestellt.

Herkömmliche Konzepte

Zu den herkömmlichen Konzepten gehören alle Methoden, welche bereits eine breite Verwendung erfahren. Die unbestritten am meisten verbreitete Methode ist auch heute noch das Bezahlen mit Bargeld. Dass sich diese Methode aber nicht als Verrechnungsmethode für Internetinhalte eignet, zeigt sich auf den ersten Blick schon daran, dass bei dieser Methode ein reales Gut, das Geld in Form von Banknoten oder Münzen, den Besitzer wechseln muss.

Die bargeldlosen Methoden, die sich im 20. Jahrhundert weit verbreitet haben und zurzeit auch als Verrechnungsmethode für Internetinhalte verwendet werden, sind vor allem die Bezahlung mit der Kreditkarte, per Lastschriftverfahren oder als Banküberweisung. Diese Methoden haben aber vor allem im Micropayment Bereich, also bei Bezahlvorgängen mit einem Gesamtvolumen von weniger als ungefähr 5 Euro, teils schwerwiegende Nachteile, die das Bedürfnis nach einer geeigneten Bezahlmethode für eben diese Volumen massiv verstärken.

Vorausbezahlte Bezahlverfahren

Der grosse Vorteil von vorausbezahlten Bezahlverfahren liegt darin, dass die Bonität des Käufers nicht gesondert ermittelt werden muss, da einzig die im voraus bezahlte Summe zur Zahlung verwendet werden kann. Grundsätzlich wird zwischen zwei Varianten vorausbezahlter Bezahlverfahren unterschieden:

1. Elektronische Geldbörse

   Als Vertreter dieser Verfahren wird im Folgenden auf die Deutsche GeldKarte [1] näher eingegangen. Mitte 1996 begannen die ersten Kreditinstitute damit, die ec-Karte für ihre Kunden mit einem Mikrochip auszustatten. Somit war die GeldKarte geboren. Heute ist die GeldKarte mit rund 63 Millionen ausgegebenen Karten das weltweit größte Geldbörsensystem.

   Sobald die Karte an einem Terminal mit maximal 200 Euro aufgeladen wurde, kann sie als Bezahlmittel eingesetzt werden. Der grosse Vorteil dieser Variante liegt in der tiefen Kommission, die zudem keinen absoluten Anteil hat. Von jeder Transaktion die mit der GeldKarte getätigt wird, fallen 0.3%, mindestens jedoch 1 Cent des Volumens als Gebühr an. Für die Bezahlung im Internet ist der Nachteil dieser Methode, dass der Kunde einen speziellen Kartenleser benötigt, der die auf der Chipkarte gespeicherten Geldwerte lesen und verifizieren kann. Der genaue Zahlungsablauf wird in der Grafik /refBildGeldkarte visuell aufgezeigt. Da diese Geräte den Heimanwender mit rund 125 Euro aber ziemlich teuer zu stehen kommen [2], bleibt abzuwarten, ob sich die GeldKarte als Verrechnungsmethode für Internetinhalte, vor allem im Micropayment Bereich, durchzusetzen vermag.

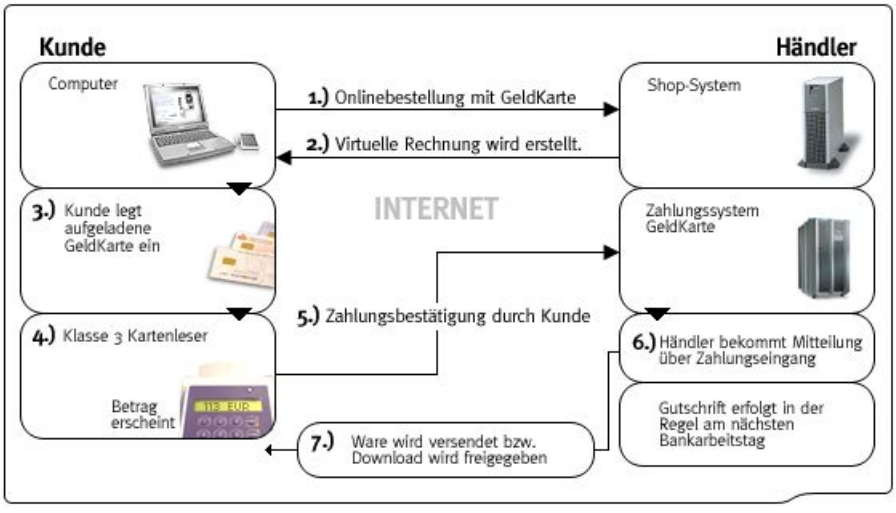

Figure 10.1: Zahlungsvorgang mit der deutschen GeldKarte

2. Scratch Card System

   Die zweite verbreitete Methode für vorausbezahlte Verrechnungsmethoden sind Rubbelkarten, so genannte Scratch Cards. Dieses System besticht durch die einfache Handhabung und die völlige Anonymität des Zahlenden. Weit verbreitet ist dieses System

bereits heute in den PrePaid Karten für Handys ohne Abonnement. Dabei bleibt zu beachten, dass diese Karten zueinander nicht kompatibel sind, zumindest nicht zur Zeit der Erstellung dieser Arbeit.

An Kiosken oder Tankstellenshops können diese Karten mit einem Wert zwischen 10 und 100 Euro bezogen werden. Unter einem Rubbelfeld findet sich ein Code, der im Internet zur Identifizierung eines Kontos mit dem der Karte aufgedruckten Kapital verwendet wird. Auf diese Weise ist sichergestellt, dass weder der Kartenanbieter noch der Verkäufer Kenntnis haben über den Käufer, obschon die Zahlung garantiert ist.

Als Vertreter dieses Systems seien an dieser Stelle die in Deutschland und Österreich verwendete Paysafecard [22], und die in der Schweiz von der Swisscom vertriebene Easyp@y Karte [14] genannt. Die bei der Verwendung dieses Systems anfallenden Kommissionen betragen bei der Paysafecard zwischen 5.5 und 19% (abhängig vom Volumen der Transaktion) und bei der Easyp@y Karte bis zu 30%.

Inkasso- / Billingsysteme

Im Gegensatz zu den vorausbezahlten Bezahlverfahren basieren die Inkasse- / Billingsysteme auf meist monatlichen Abrechnungen. Dabei wird grundsätzlich unterschieden zwischen einer Rechnungsbeziehung mit dem bereits vorhandenen Internet Service Provider oder einem Drittanbieter:

1. Rechnungsbeziehung mit Internet Service Provider

   Der grosse Vorteil eine Rechnungsbeziehung mit einem Internet Service Provider liegt in dem bereits vorhandenen Vertragsverhältnis. Der Kunde braucht sich je nach verwendetem System auch nicht extra für jede Transaktion anzumelden, da er über seinen Provider ja bereits identifiziert ist. Die Anonymität ist, von einer Verschwiegenheit des Providers ausgehend, gewahrt. Grundsätzlich ist der Internet Service Provider selbst einem Verbund mehrerer Provider für internetbasierte Verrechnungsmethoden angeschlossen, oder er benutzt unter Lizenz ein System eines externen Anbieters.

   Bill-it-easy [3] ist ein Beispiel für ein System, das mehrere Internet Service Provider zu einer Zahlungsform vereint. Das von Swisscom verwendete Zahlungssystem click&buy [4] ist hingegen Teil einer internationalen Allianz, unter anderem bestehend aus der British Telecom in England, der FIRSTGATE Internet AG in Deutschland und Amerika, sowie der Österreichischen Raiffeisenbank Gruppe. Dieses Zahlungssystem existiert ebenso als Rechnungsbeziehung mit einem Drittanbieter, worauf im Folgenden eingegangen wird.

2. Rechnungsbeziehung mit Drittanbieter

   Die bereits erwähnte click&buy Zahlungsmethode existiert in vielen Ländern auch als Rechnungsbeziehung mit einem Drittanbieter - meist ist dies die FIRSTGATE Internet Gruppe [5]. Dem Nachteil einer neuen Vertragsbeziehung steht der Vorteil

einer getrennten, und damit besser anonymisierten Rechnung gegenüber. Die Kommissionen für eine Transaktion betragen zwischen 7 und 35 Prozent des Volumens.

Vor allem seit 2002 von eBay, einem der weltweit grössten Online-Marktplätze [6] übernommen, existiert mit PayPal [7] eine weitere, mit 56 Millionen Konten weltweit sehr verbreitete Zahlungsmethode. Der grosse Vorteil bei dieser Methode ist die spesenfreie Überweisung von Geldbeträgen von einem PayPal Konto auf ein anderes. Grafik /refBildPayPal1 stellt eine Überweisung des Käufers an den Verkäufer dar, wobei PayPal als zwischengeschaltete Instanz fungiert. Einzig für die Bezahlung von Internetinhalten werden Kommissionen von 1.9 bis 3.4% des Volumens zuzüglich 35 Cent pro Transaktion verrechnet.

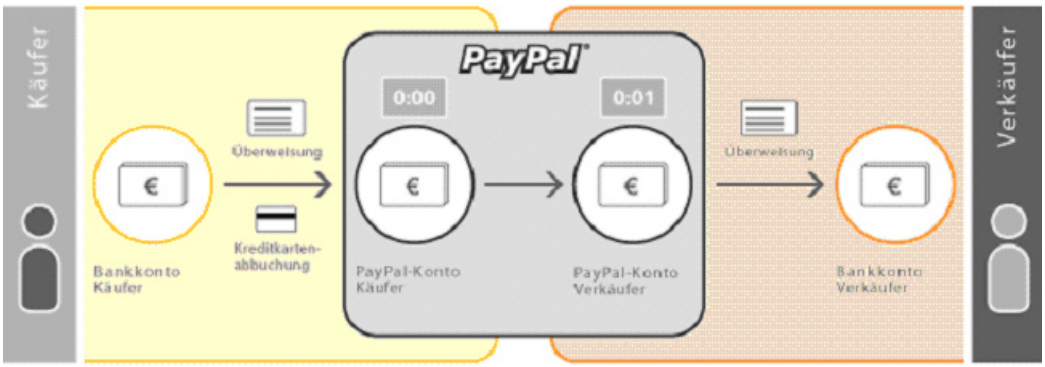

Figure 10.2: Zahlungsvorgang mit PayPal

Mobile Payment Systeme

Da mittlerweile das Handy in den meisten industrialisierten Ländern zu einem völlig normalen Alltagsgegenstand geworden ist, erstaunt es auch nicht, dass mobile Bezahlsysteme nicht lange auf sich warten liessen.

Bereits 1999 wurde mit der paybox.net AG eine Gesellschaft für mobiles Bezahlen mit dem Telefon gegründet [8]. Der Bezahlvorgang läuft bei dieser Methode wie folgt ab. Der Käufer übermittelt dem Verkäufer die Bestelldaten, darunter auch seine Handynummer, beziehungsweise seine Kundennummer bei paybox. Anschliessend liefert der Verkäufer die zur Verifikation der Transaktion notwendigen Daten an paybox weiter. Diese wiederum verifiziert die Transaktion mit einem Rückruf beim Käufer. Erst danach erfolgt die Lieferung und anschliessende Lastschrift beim Käufer und Überweisung an den Verkäufer. Dieser Transaktionsvorgang wird in Grafik /refBildHandyPaybox schematisch dargestellt.

Dieses System schien sich aber nicht überall durchzusetzen. So wurde paybox im Jahre 2003 in Deutschland erst von der Firma Moxmo übernommen und nach verschiedenen Problemen [9] eingestellt. In Österreich wurde paybox im Jahre 2003 hingegen von der mobilkom austria AG & Co KG übernommen und unter gleichem Namen weitergeführt.

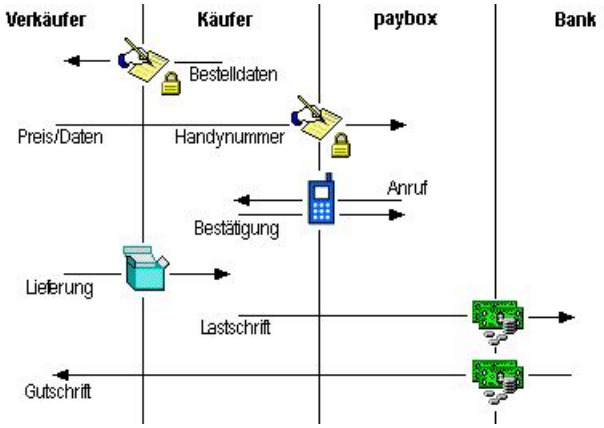

Figure 10.3: Zahlungsvorgang mit Paybox

So ist paybox auch heute in Österreich aktiv und ermöglicht nebst Internetzahlungen auch Bezahlungen bei mobilen Dienstleistern (z.B. Taxi) oder im stationären Handel (z.B. Hotel). Ein grosser Unterschied bei diesem System ist die Jahresgebühr von 15 Euro für jeden paybox Benutzer. Die Onlinehändler haben bei der Verwendung von paybox eine Transaktionsgebühr von 2%, mindestens jedoch 50 Cent zu entrichten.

Eine weitere Möglichkeit der mobilen Bezahlsysteme liegt in den sogenannten Premium SMS. Da Kurzmitteilungen unterschiedlich hoch taxiert werden können, eignen sie sich durchaus für einfache Transaktionen im Micropayment Bereich. Mit solchen SMS sind aber oft auch unseriöse Unternehmen verbunden sind, da mit nur einem SMS rechtlich gesehen ein Vertrag zustande kommen kann [10].

10.1.3 Vergleich vorgestellter Systeme

An dieser Stelle sollen die vorgestellten Verrechnungsmethoden für Internetinhalte in einer kurzen Übersicht miteinander verglichen werden. In der folgenden Übersicht /refBildUebersichMethoden sind die aus Autorensicht jeweils grössten Vor- und Nachteile der einzelnen Methoden aufgelistet.

Konzept	Vorteil	Nachteil
GeldKarte	Tiefe Kommissionen (0.3%)	Aufwendige Installation
Paysafecard	Anonym (wenn real gekauft)	Wenig verbreitet
FirstGate	Verbreitet, (CH: Swisscom als Partner)	Hohe Kommissionen (7-35%)
PayPal	Weit verbreitet, E-Bay als Träger	Min. 0.35€ Kommission
Paybox	Handy ist Alltagsgegenstand	Jahresgebühr für Käufer

Figure 10.4: Übersicht der Vor- und Nachteile der vorgestellten Methoden

Wird der Tatsache Beachtung geschenkt, dass von den verglichenen Methoden vor allem diejenigen mit einem starken und seriösen Partner bzw. Inhaber teils grossen Zuwachs

vermelden können (click&buy von Swisscom oder PayPal von eBay), so lässt sich bei diesem Vergleich ein Schluss deutlich ziehen: Um als Verrechnungsmethode für Internetinhalte vor allem international erfolgreich zu sein, ist ein grosser wirtschaftlicher Rückhalt unabdingbar. Obwohl ein relativ neues und durch die Online Umgebung sehr technisches Gebiet also definitiv keine Option für aufkommende .Net Unternehmen.

10.2 Wieso haben sich viele Methoden nicht oder nur lokal durchgesetzt

> In der Phase des Internet-Hypes prognostizierten führende private Forschungsinstitute bis 2005 jährlich mehr als eine Verdopplung des eBusiness-Umsatzes. Doch diese exorbitanten Erwartungen wurden enttäuscht, das Volumen entwickelte sich deutlich gemäßigter. Eine Ursache liegt darin, dass die Pioniere des eBusiness die Schwierigkeiten eher bei der Produktpalette, dem Marketing und der Logistik sahen. Dabei unterschätzten sie zunächst die Herausforderungen des nachgelagerten Bezahlvorgangs.[11]

Dies deutet schon daraufhin, dass der Bezahlvorgang beim Online-Shopping ein kritischer Erfolgsfaktor ist. Dieser Abschnitt beschreibt die aktuellen Anforderungen und Probleme aus Sicht der Kunden, der Händler, sowie allgemeneiner Gesichtspunkten.

10.2.1 Charakteristika von Benutzern und Methoden

Die Marktposition von Zahlungssystemen ist von verschiedenen Faktoren abhängig, welche von verschiedenen Gruppen geprägt werden. Welche und wie diese Faktoren von der Kundenseite im Online-Handel beeinflusst werden, wird im Folgenden beschrieben.

Verbreitung und Beliebtheit von Zahlungssystemen

Bei Studien über die Verbreitung von verschiedenen Zahlungsmethoden für Internetinhalte hat sich eine klare Neigung zu den klassischen Systemen ergeben wie in der Grafik 10.5 klar zu erkennen ist. Die Aussicht auf 12 Monate weist weiter darauf hin, dass sich diese Neigung nicht so bald ändern wird, die klassischen Zahlungsmethoden sogar noch zulegen werden.

Die Erklärung hierfür liegt hauptsächlich bei den Endkunden. Die beliebtesten Zahlungsmethoden, sind natürlich die, die die geforderten Eigenschaften möglichst beinhalten. Diese Eigenschaften sind, wie man in untenstehender Grafik sieht, vorallem Sicherheit und Einsatzfähigkeit.

`Erst die Ware, dann das Geld` ist nach wie vor die sicherste Methode beim Einkauf. Und dies trifft ja zumindest auf einen Teil der üblichen Zahlungsmethoden zu: sowohl auf Bezahlung per Nachname als auch per Rechnung. Für den Kunden ideal, für die Branche

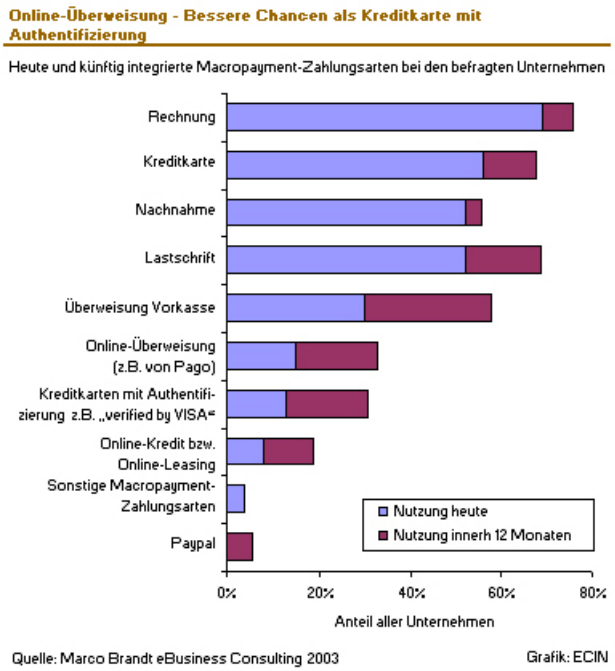

Figure 10.5: Verbreitung von Online-Zahlungsmethoden

bedeutend schwieriger. Darauf wird jedoch erst im zweiten Teil "'allgemeine Probleme von Online-Zahlungsmethoden"' eingegangen.

Ausnahmen machen dabei nur die Lastschrift und die Kreditkarte. Diese Methoden verweisen direkt auf den zweiten Punkt in der Grafik 10.6 dem `Vertrauen in den Abwickler der Transaktion`: vor Allem das Lastschriftverfahren wird im Allgemeinen nur eingesetzt, wenn der Anbieter als sehr Vertrauenswürdig eingestuft wird.

Mit der Kreditkarte hingegen wird des öfteren freizügiger umgegangen, was im Anbetracht der Betrügereien rund um diese Zahlungsmethode im ersten Moment überraschen mag. Doch die rechtliche Lage ist voll auf den Kunden ausgelegt, Kosten durch Betrügereien landen im Allgemeinen beim Händler oder der Bank.

Eine weitere wichtige Anforderung - und der klarste Grund für die Benutzung klassischer Methoden - ist die Einsatzfähigkeit und die Benutzerfreundlichkeit: Bezahlvorgänge per Rechnung sind heute Alltag, sie sind in fast allen allen Branchen der Wirtschaft vertreten und benötigt somit nichts Unbekanntes. Dies gilt analog für die Kreditkarte oder die Bezahlung per Nachnahme.

Als Essenz daraus lässt sich sagen, dass vorallem Bequemlichkeit und mangelndes Vertauen User von neuartigen Zahlungsmethoden abhält.

10.2.2 allgemeine Probleme von Online-Zahlungsmethoden

Wenn denn nun die klassischen Methoden auf solche Beliebtheit stossen, wieso wird dann so fieberhaft nach neuen Möglichkeiten gesucht? Das Problem liegt bei den anderen Parteien eines Handels: den Händlern und den Banken.

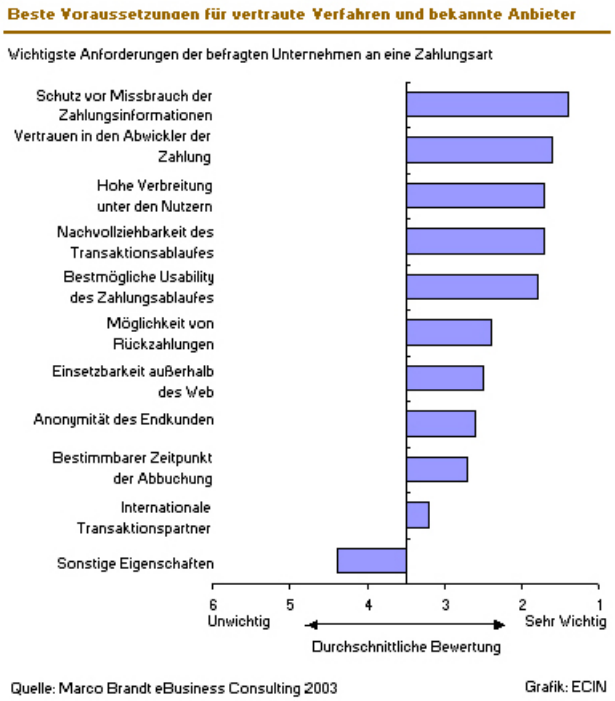

Figure 10.6: Anforderungen an Zahlungsmethoden

Wieso in die Ferne schweifen...

Sowohl bei Rechnung, Lastschrift, Nachnahme als auch Kreditkarte liegt das größere Risiko beim Händler. Und dies sind nun mal die am häufigsten (...) angebotenen Zahlungsmittel. So wird zwar beispielsweise von über der Hälfte der Online-Shops die Zahlung per Lastschrift angeboten, indem der Kunde lediglich seine Kontonummer und Bankleitzahl mitteilt, doch ist das Missbrauchsrisiko, ebenso wie bei der Kreditkartenzahlung, ohne Unterschrift sehr hoch.

Der vermeintliche Kunde kann sowohl Lastschrift- als auch Kreditkartenzahlungen problemlos wieder stornieren. Fehlt die Unterschrift beim Lastschriftverfahren, hat der Kunde das Recht, nahezu beliebig lange dem Lastschrifteinzug bei seiner Bank zu widersprechen. Aber auch bei vorliegender Unterschrift besitzt der Kunde immerhin noch ein sechswöchiges Widerspruchsrecht.

Ganz ähnlich verhält es sich bei Kreditkartentransaktionen im Internet. Hier gelten die gleichen Bedingungen wie beim Versandgeschäft (Mailorder). Der Verkäufer hat keine Zahlungsgarantie und obendrein die Nachweispflicht für die korrekte Warenanlieferung beim Kunden. Ohne Unterschrift übertragen die Kreditkarten-Organisationen dem Händler das volle Risiko bei einer Reklamation. Sowohl bei Lastschrift als auch bei Kreditkarte bleibt der Händler bei Beanstandungen in jedem Fall auf den so genannten Chargeback-Gebühren sitzen. So kommt es bei Online-Händlern allzu oft vor, dass der Rechnungsbetrag entweder erst gar nicht abgebucht werden kann oder vom Kunden wieder zurückgebucht wird. In der Regel hat die Ware dann jedoch das Lager schon längst verlassen.

> Die Risiken der Kreditkartenzahlung für den Händler verdeutlicht auch eine aktuelle Einschätzung von Eurocard. Demnach büßen Online-Händler bis zu zehn Prozent ihres Jahresumsatzes durch Kreditkartenbetrug ein. Die Zuwachsrate bei den betrügerischen Online-Geschäften liegt laut Eurocard-Geschäftsführer Manfred Krüger bei 40 Prozent.[12]

Figure 10.7: Belastungszeitpunkt verschiedener Zahlungsmethoden

In der Grafik 10.7 wird das Problem beider Seiten nochmals dargestellt. Auf Seiten der Kunden ist das Ausfallen von Zinsen natürlich im Allgemeinen das kleinste Übel, bei Vorkasse ist meistens die Unsicherheit Grund für die ablehnende Haltung.

Folglich ist trotz der Beliebtheit bei den Kunden die aktuelle Situation für den Handel nicht befriedigend und die Suche nach neuartigen Methoden gerechtfertigt. Dabei tritt jedoch ein grundlegendes Problem auf:

ein Henne-Ei-Problem

Ohne die Akzeptanz der Kunden wird sich eine neuartige Zahlungsmethode nie durchsetzen. Grundsätzlich setzen sich auf dem Markt vor allem einfache und intuitive Lösungen durch. Alles, was Hard- und/oder Softwareinstallationen benötigt oder nur schon komplizierte Registrierungen benötigt stösst schnell auf Widerstand.

Aus diesem Dilemma ergibt sich das, was hier als `Grounding-Kausalkette für neue Zahlungsmethoden` bezeichnet wird:

> Eine neuartige Zahlungsmethode, die vom Prinzip her gut funktionieren würde, wird von einem einzelnen Unternehmen entwickelt und eingesetzt. Die Technik ist jedoch noch nicht genug ausgereift, Sicherheitsmängel und Probleme in der Benutzerfreundlichkeit lassen die Kunden abschrecken. Andere Online-Händler werden nicht auf die neue Methode aufmerksam, das entwickelnde Unternehmen muss die Bemühungen wieder einstellen.

Sicherlich ist dies zu verallgemeinernd formuliert, jedoch hat es schon eine Kernaussage, die auf die meisten gescheiterten Entwicklungen in diesem Bereich zutrifft.

Viele Anbieder versuchen mit den unterschiedlichsten Methoden ein Stück vom Online-Handel-Kuchen abzubekommen. Doch die fehlenden Standards aufgrund fehlender Zusammenarbeit sind umgehend wieder mitverantwortlich dafür, dass sich Kunden Händler suchen, bei dem sie Zahlungsmethoden anwenden können, die sie kennen und in weiter Verbreitung anwenden können. Es ist also wichtig, dass ausgewachsene Lösungen möglichst breit eingeführt werden, um so die Akzeptanz und das Vertrauen der Kunden in neuartige Zahlungssysteme zu erlangen.

Micropayment - eine zusätzliche Herausforderung

Ein zusätzlicher Ansporn, neue Zahlungsmethoden für das Internet bereitzustellen, ist die Problematik des Micropayment. Die klassischen Methoden sind für Bezahlungen von Beträgen bis 5 Euro nicht geeignet. Das Problem der Gebühren am Beispiel der Kreditkarte verdeutlicht das.

Immerhin werden im Internet (Versandgeschäft) von den Kreditkarten-Organisationen Gebühren zwischen 3 und 5 Prozent verlangt, plus zusätzlicher Kosten pro Transaktion, Einrichtungsgebühren (setup fee) und monatliche Mindestgebühren.

Ähnlich verhält es sich auch bei Überweisungen speziell ins Ausland, wo hohe Transaktionskosten anfallen. Diese werden zwar gerne an die Kunden abgewälzt, was diese aber natürlich auch nicht auf sich nehmen wollen.

Somit gibt es zur Zeit keine weit verbreitete standardisierte Lösung für das Micropayment-Problem, was neue Entwicklungsanstrengungen rechtfertigt.

10.2.3 Scheitern neuartiger Zahlungssysteme an einigen spezifischen Beispielen

Schon viele visionäre Anstrengungen wurden unternommen, um neue Zahlungsmethoden für das Internet zu verbreiten. Jedoch sind die meisten davon nach mehr oder weniger kurzer Zeit auch schon wieder vom Markt verschwunden. Im Folgenden werden einige dieser gescheiterten Projekte betrachtet, aber auch ein Beispiel, wie es funktionieren kann dargestellt.

eCash

Die Übertragung des Bargeldmodells auf das Internet war durchaus eine grosse Aufgabe. Geschafft hat dies David Chaum, durch den Einsatz komplexer Verschlüsselungsmethoden auf mathematischer Ebene. Das Problem dabei, es konnten sich nur Intellektuelle für die Anonymisierung des Cybermünzen faszinieren. Eine geniale Idee, technisch umsetzbar, aber ohne Echo in der Wirtschaft.

Dennoch hat dieser Versuch gezeigt, dass es zumindest technisch machbar wäre und wird vielleicht irgendwann wieder aufgegriffen und markttauglich präsentiert.

CyberCash

CyberCash hatte einen tollen Namen, mit dem viele sogar die gesamte Szene identifizieren. Aber die Kunden sind nicht bereit, sich extra eine neue Software herunterzuladen, wenn dahinter nicht ein richtiger Nutzen steht, denn mit CyberCash konnte so gut wie nichts erworben werden. Folglich hätte noch viel investiert werden müssen, um die CyberCoins in die Hände von mehr Händlern und Verbrauchern zu bringen. (in Anlehnung an [13])

Da haben die verantwortlichen Banken entschieden, dass ihnen das Ganze als Demonstrationsobjekt nicht so wichtig ist. Zumal in Deutschland hohe Lizenzgebühren gegenüber der amerikanischen Mutterfirma anfielen, die nicht durch entsprechende Gewinnaussichten kompensiert wurden.

SET - ein Gegenbeispiel

SET (Secure Electronic Transaction) lebt noch und wird weiter entwickelt. Was macht den Unterschied zu eCash oder CyberCash aus? Wäre SET wie CyberCash nur in der Hand eines einzigen Anbieters, wäre es wohl schon genauso vom Markt verschwunden. Aber so wird in dem Konsortium weiter versucht, SET weiterzuentwickeln. Man hat gemerkt, dass benötigte Zusatzsoftware bei den Kunden unbeliebt ist, so entstand die Idee, SET über den Server des Anbieters laufen zu lassen.

Der Erfolg ist noch nicht garantiert, und der Weg dorthin sicher nicht einfach. Dennoch kommen schon neue Produktangebote für diesen Bereich auf den Markt. Somit wäre für die Bequemlichkeit und zumindest ansatzweise für Verbreitung gesort, abzuwarten ist nur noch die Akzeptanz der User.

10.3 Marktchancen der Methoden

> Elektronische Zahlungsmethoden müssen genauso wie herkömmliche Zahlungsmethoden allen beteiligten (üblicherweise zwei) Transaktionspartner die Sicherheit bieten, dass die Transaktion wie erwartet stattfindet, und der Wert des Zahlungsbetrags rechtmässig aus dem Zugriffsbereich des Zahlenden in den Zugriffsbereich des Zahlungsempfängers transferiert wird. Gleichzeitig soll die Zahlung schnell und intuitiv ausgelöst und abgewickelt werden können. Ausserdem sollten die Transaktionskosten im Vergleich zum Zahlungsbetrag gering sein.
>
> Während sich im Macropayment-Bereich verschiedene Zahlungsverfahren etabliert haben, gibt es im Micropayment-Bereich bisher erst verschiedene an sich vielversprechende Lösungsansätze, von denen sich bisher allerdings noch keiner dauer-

> haft durchsetzen konnte, obwohl sie den genannten Forderungen grundsätzlich entsprochen haben.
>
> Da Netzwerkgut ... Damit sich eine Zahlungsmethode durchsetzen kann, muss diese nämlich auch bei einer grossen Anzahl Anbieter eingesetzt werden können, und ebenso sollte dieses Zahlungsverfahren von einer möglichst grosse Anzahl Benutzer eingesetzt werden.

10.3.1 Anforderungen an eine Zahlungsmethode

Als Gradmesser für die Tauglichkeit einer Zahlungsmethode sollen die an sie gestellten Anforderungen dienen, welche an folgender Stelle erarbeitet werden.

Erfolgskritische Faktoren

Verschiedene Faktoren beeinflussen die Attraktivität einer Zahlungsmethode im Internet und entscheiden damit auch darüber, ob sich eine Zahlungsmethode durchsetzen kann.

1. Akzeptanz durch eine grosse Anwendergruppe
2. Akzeptanz durch eine grosse Anbietergruppe (auch ausserhalb des Internet)
3. Identität des Anbieters der Zahlungsmethode
4. Sicherheit und Anonymität der Zahlungstransaktionen
5. Einfache Handhabung
6. Transaktionskosten nahe bei Null

(In Anlehnung an [25] [16])

Bedeutung der weiten Verbreitung (Käufer)

Zahlungsmittel und -verfahren (auch elektronisches) kann zu den Netzwerkgütern gezählt werden, deren Wert umso grösser ist, je mehr Benutzer es gibt. Eine grosse Zahl Benutzer steigert das Vertrauen in eine Zahlungsmethode (”‘unzählige zufriedene Nutzer können nicht irren”’). Ausserdem steigt mit der Anzahl Benutzer die Wahrscheinlichkeit, dass dass bei positiven Erfahrungen eines Kunden eine Mund zu Mund Propaganda in dessen Umfeld durchaus für eine weitere Verbreitung sorgen kann.

Schliesslich ist es auch für die Anbieter vor allem dann interessant, eine Zahlungsmethode anzubieten, wenn es auch genügend Kunden gibt, die diese Methode nutzen möchten. [18] Nur so lohnt sich für einen Anbieter die Investition, die für Anpassung oder Integration der Zahlungsmethode in die bestehende Informatik-Infrastruktur nötig ist.

Bedeutung der weiten Verbreitung (Anbieter)

Genau so ist es auch für die Anwender vor allem dann interessant, eine elektronisches Zahlungsmittel zu benützen, wenn es bei einer grossen Zahl von Anbietern eingesetzt werden kann. Einerseits wächst die empfundene Sicherheit, wenn eine Zahlungsmethode von einer Vielzahl von Anbietern akzeptiert wird. Ausserdem steigt mit der Anzahl Anbieter auch die Wahrscheinlichkeit, dass ich eine bestimmte Zahlungsmethode nutzen muss, wenn ich bei der Nutzung eine positive Erfahrung gemacht habe, werde ich dieses Verfahren voraussichtlich auch ein weiteres mal nutzen.

Weiter wird der sich der potentielle Nutzer eines Zahlungsverfahrens, für das man sich vor der Nutzung registrieren muss, eher die Zeit und Mühe für die Registrierung nehmen, wenn er dieses Benutzerkonto auch bei anderen Anbietern nützen kann. [18]

Bedeutung von Real-World Anbietern

Kann eine neue elektronische Zahlungsmethode auch bei Anbietern aus der realen Welt angewandt werden, so kann das gewisse Eintrittsbarrieren herabsetzen. Einerseits kann ein neuer Nutzer bei allfälligen Probleme direkt Unterstützung durch den Verkäufer erhalten. Andererseits kann es das Vertrauen in ein Verfahren erhöhen, denn sollten bei einer Zahlung Probleme auftauchen, kann ich mich möglicherweise beim entsprechenden Anbieter vor Ort erkundigen oder beschweren. Schliesslich wird auch die Zahl von potentiellen Anbietern erhöht, wenn mit der Zahlungsmethode auch bei Real-World Anbietern eingesetzt werden kann.

10.3.2 Macropayment-Bereich

Der Macropayment-Bereich beginnt, wie bereits weiter oben erwähnt, ab einem Zahlungsvolumen von zirka 10 CHF. Damit ergibt sich eine grundlegend andere Ausgangslage bei der Marktbeurteilung als im Micropayment-Bereich.

Ausreichend etablierte Zahlungsverfahren

Wie im vorangegangenen Kapitel 2 betrachtet gibt es im Macropayment-Bereich bereits heute unterschiedliche etablierte Methoden. Diese Zahlungsmethoden erfüllen zumindest die meisten Ansprüche an eine Zahlungsmethode, beispielsweise Sicherheit und Anonymität, dem Zahlungsbetrag angemessene Kosten, oder aber auch einfache Handhabung, ausreichend. Diese Zahlungsmethoden sind ausserdem genauso von vielen Anbietern wie auch von Käufern akzeptiert. Die etablierten Methoden werden deshalb vermutlich in näherer und mittlerer Zukunft nicht wesentlich ändern.

Electronic Bill Presentment and Payment (EBPP)

Eine der wenigen Neuerungen in diesem Bereich ist das Electronic Bill Presentment and Payment (EBPP).

> Mit Electronic Bill Presentment and Payment (EBPP) wird die elektronische Rechnungsstellung und elektronische Bezahlung von Rechnungen bezeichnet. Die elektronische Abwicklung des Rechnungsprozesses durch EBPP beinhaltet zwei Teilprozesse. Zum einen ist dieses das Zahlen (Electronic Bill Payment) und zum anderen das Präsentieren und Versenden von Rechnungen (Electronic Bill Presentment). [24]

Das elektronische Bezahlen von Rechnungen ist bereits heute recht weit verbreitet, viele sparen sich schon heute den Gang zur Bank oder Post, und begleichen ihre Rechnungen, die sie üblicherweise per Briefpost oder gleichzeitig mit der einer Warenlieferung zugestellt erhalten haben, per e-Banking von zu Hause aus. Abgesehen davon, dass die Finanzinstitute so Aufwand für Geldtransporte und Schalterpersonal einsparen können, können elektronisch erfasste Zahlungsaufträge auch direkt weiterverarbeitet werden, während manuelle Zahlungen erst noch elektronisch erfasst werden müssen.

Mit EBPP soll zusätzlich der Aufwand für Rechnungsausdruck, -verpackung und -versand, sowie Portokosten eingespart werden, in dem die Rechnungen nicht mehr auf Papier, sondern nur noch elektronisch zugestellt werden. Idealerweise kann der Zahler die elektronisch zugestellte Rechnung mit wenigen Mausklicks bezahlen, ohne dass er sämtliche Daten in einem neuen Formular abtippen muss. In Grafik /refBildEBPP wird diese Zwischenschaltung eines Finanzinstitutes aufgezeigt.

Figure 10.8: Übersicht Electronic Bill Payment

Nachdem EBPP im Zahlungsverkehr zwischen Unternehmen im Zusammenhang mit EDI schon länger eingesetzt wird, werden nun auch für den Zahlungsverkehr zwischen Unternehmen und Privatpersonen EBPP Systeme lanciert. In der Schweiz gibt es mittlerweile zwei Systeme, die auf dem EBPP Prinzip beruhen: yellowbill (Postfinance) [Yellowbill] und PayNet (verschieden Kantonalbanken, Credit Suisse, UBS und andere) [Paynet]. Dass sich die verschiedenen Finanzinstitute nicht auf einen einheitlichen Standard einigen konnten, und sich damit die Anbieter- und die Benutzerzahl auf verschiedene Zahlungsverfahren aufteilt, stellt eines der wenigen Probleme dieser Zahlungsverfahren dar. Ausserdem sind die Zahlungsmethoden jeweils nicht global einsetzbar, was damit im Zusammenhang steht, dass auch die herkömmlichen Zahlungsverfahren, an welche sich die elektronischen anlehnen, national genormt sind.

Da sich die neuen, elektronischen Zahlungsverfahren sehr stark an bekannte Zahlungsverfahren anlehnen, und diese bekannten Verfahren erleichtern, da ausserdem eine grössere Anzahl von seriösen Partnern die neuen Zahlungsmethoden mittragen, stehen die Chancen gut, dass sich diese durchsetzen und etablieren.

10.3.3 Micropayment-Bereich

Zahlungen im Micropayment-Bereich, deren Volumen sich in einem Bereich von bis zu etwa 10 CHF bewegen, haben aufgrund eben dieser Transaktionsvolumen eine eigene Rolle im Bereich der Verrechnungsmethoden.

Überblick über die Varianten

Wie in Kapitel 2 angesprochen besteht das Hauptproblem im Micropayment-Bereich darin, dass die Kosten für eine Zahlung nach herkömmlichen Verfahren (mit Kreditkarte, Rechnung, Nachnahme, etc.) im Verhältnis zum Zahlungsbetrag zu hoch sind. Da es nur schwer möglich ist, die Kosten für eine Zahlung zu senken, lassen sich die die relativen Kosten einer Transaktion nur dadurch senken, dass mehrere Transaktionen zusammengefasst und gemeinsam verrechnet werden. Wird diese kumulierte Zahlung fällig, bevor die eigentliche Micropayment-Transaktion stattfindet, spricht man von Prepay Systemen, werden die Micropayment-Transaktionen gesammelt und zu einem späteren Zeitpunkt zusammen verrechnet, spricht man von Postpay oder auch Billing Systemen.

Auswirkungen auf eBusiness

Bis vor kurzem hat eine Mehrheit der Anbieter von digitalen Inhalten wie beispielsweise Zeitungen und Zeitschriften diese gratis zur Verfügung gestellt, nicht zuletzt deshalb, weil es für die Verrechnung keine kostendeckenden Zahlungsverfahren gab. Finanziert wurde dieses Angebot bisher meist ausschliesslich über Werbung, obschon bei den Nutzern eine wachsende Zahlungsbereitschaft vorhanden ist wie der Grafik /refBildZahlungsbereitschaft zu entnehmen ist (Quelle: [17]).

Mehr und mehr werden diese Angebote nun aber zahlungspflichtig, sei es, dass interessierte Nutzer ein spezielles Online-Abonnement lösen müssen, oder sei es, dass Artikel nur gegen Bezahlung mit einer der nachfolgend vorgestellten Micropayment Zahlungsmethode erhältlich sind. Die Einführung solcher Zahlungsverfahren eröffnet also auch neue Geschäftsfelder und macht die neuen Zahlungsverfahren damit auch wirtschaftlich interessant.

10.3.4 Prepay Systeme

Der grosse Vorteil von Prepay Systemen liegt in der garantierten Zahlung, da der Betrag bereits vorfinanziert ist. Eine gesonderte Bonitätsprüfung des Schuldners ist also nicht mehr notwendig.

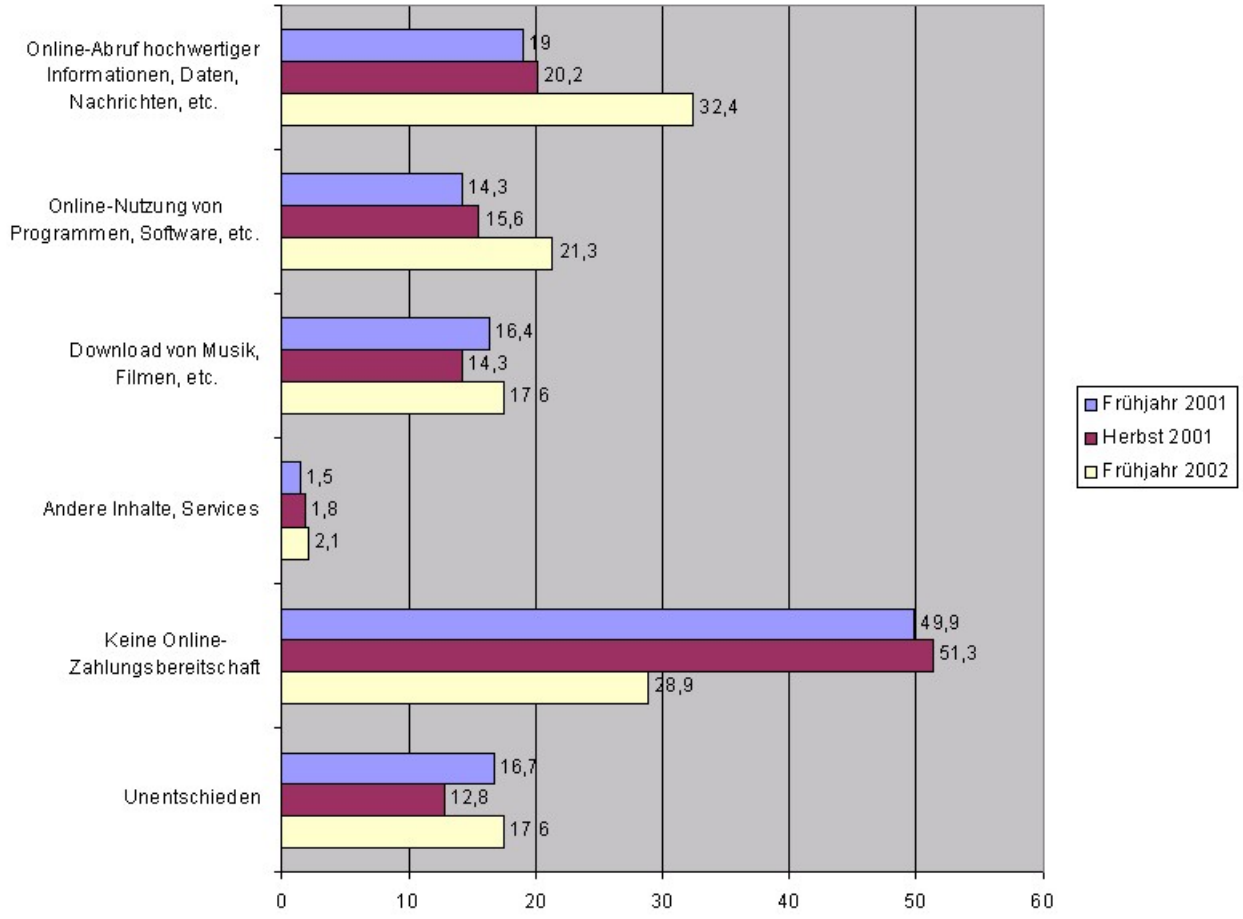

Figure 10.9: Übersicht online Zahlungsbereitschaften

Elektronische Währungen

Die intuitivste Variante für ein elektronisches Zahlungssystem wäre vermutlich eine elektronische Währung mit elektronischen Geldstücken und / oder -noten, die man wie in der realen Welt üblich zu einem Wechselkurs in andere (reale) Währungen umtauschen kann. Diese Variante konnte auch von verschiedenen Anbietern technisch realisiert werden. Zu den Beispiele für elektronische Währungen, die aber gescheitert sind, gehören eCash oder CyberCash. Gründe für das Scheitern wurden in Kapitel 2 gezeigt.

Die Idee von virtuellen Währungen ist bestimmt noch nicht komplett vom Tisch, möglicherweise war die Zeit noch nicht reif dafür. Es gibt genügend Beispiele dafür, dass sich gute Ideen manchmal erst beim zweiten oder dritten Anlauf durchzusetzen vermögen. Vielleicht wird es in Zukunft sogar Aufgabe des Staates (resp. der staatlichen Notenbank) sein, neben einer realen Währung auch eine virtuelle Währung herauszugeben.

Prepay Karten

Prepaid Karten lassen sich in zwei Kategorien einteilen: Einerseits gibt es Wertkarten, die man gegen Bargeld an Poststellen, Kiosken oder Tankstellen beziehen kann, andererseits gibt es chipbasierten Karten, die man an Bankomaten oder auch über Kreditkartenbelastung im Internet aufladen kann.

Bei Wertkarten ist üblicherweise eine mehrstellige Kartennummer aufgedruckt, die nach dem Kauf freigerubbelt und bei Einkäufen im Internet eingegeben muss, bis der Wert der Karte aufgebraucht ist. Beispiele für solche Wertkartensysteme sind Swisscom Easyp@y (Schweiz) [14], Micromoney von der Deutschen Telekom (Deutschland) [19] oder die paysafecard (Österreich, Deutschland) [22]. Diese Wertkartensysteme beruhen auf einer sehr einfachen Technik, sind sehr intuitiv in der Anwendung und die Gebühren sind sowohl für die Anwender wie auch für die Anbieter sehr günstig. Hinter den meisten dieser Produkte steht zudem häufig ein namhafter und vertrauenswürdiger Anbieter (z. B. das nationale Telekom Unternehmen). Die Chancen stehen deshalb gut, dass sich diese Wertkarten etablieren können.

Im Falle von chipbasierten Karten muss der Anwender üblicherweise ein Kartenlesegerät besitzen, deren Anschaffung nicht günstig ist, stehen die Chancen eher schlecht, dass sich diese Prepaid Karten etablieren können.

PayPal

Das PayPal Zahlungsverfahren wurde 1998 von der Firma PayPal international entwickelt. Richtig durchsetzen konnte sich dieses Verfahren allerdings erst, nachdem PayPal international im Jahr 2002 vom Internet Auktionshaus Ebay übernommen wurde. Es gibt heute einen riesigen, ca. 56 Millionen PayPal Kontoinhaber in 45 Ländern umfassenden Benutzerkreis. [21] Damit ein Käufer PayPal nutzen kann, müssen er und der Verkäufer Inhaber eines PayPal-Kontos sein. Tätigt der Käufer eine Zahlung, indem er sich bei PayPal einloggt, die Email-Adresse des Käufers und den Betrag eingibt, der überwiesen werden soll. Der Verkäufer erhalt dann eine Email mit dem Hinweis, dass seinem Konto Geld gutgeschrieben wurde ("'You've got Money"'). Im Email ist ein Link angegeben, wo der Verkäufer den überwiesenen Betrag überprüfen kann. Der überwiesene Betrag wird schliesslich dem Konto oder der Kreditkarte des Käufers belastet. Guthaben auf dem PayPal Konto kann der Kontoinhaber seinem Bankkonto gutschreiben lassen. PayPal ist somit nicht eine reine Prepaid Zahlungsmethode, denn das Bankkonto oder die Kreditkarte des Käufers wird erst im Nachhinein belastet. Da aber für die Guthaben auf dem PayPal Konto keine Zinsen anfallen, kann man diese Zahlungsmethode trotzdem auch zu den Prepaid Methoden zählen.

PayPal ist für die Nutzer im Wesentlichen kostenlos, was diese Zahlungsmethode sehr attraktiv macht. PayPal finanziert sich hauptsächlich über die Zinsen, welche die Guthaben auf den Konten der PayPal-Nutzer abwerfen. [15] Die günstigen Nutzungsgebühren, genauso wie die intuitive Bedienung haben dazu beigetragen, dass PayPal so weit verbreitet ist. Auch wenn es verschiedene Kritiker von PayPal gibt, wird das Zahlungsver-

fahren wohl weiterhin sehr beliebt sein. Grafik /refBildPayPal2 zeigt schematisch einen typischen Zahlungsvorgang unter Verwendung von PayPal auf.

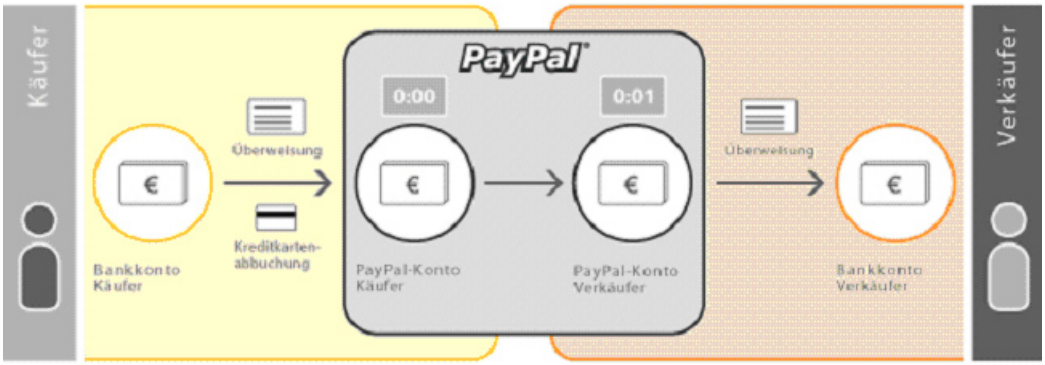

Figure 10.10: Übersicht PayPal

10.3.5 Billing Systeme

Bei den am meisten verbreiteten Billing (oder auch Postpaid genannten) Systemen werden die Micropayment Transaktionen gleichzeitig mit einer anderen Rechnung abgerechnet, üblicherweise sind das Rechnungen von Telekom-Anbietern. Dies deshalb, weil die Authentifizierung des Kunden in den meisten Fällen auch über Festnetz- oder Mobiltelefon geschieht.

Mobile Payment mPayment

Bei mPayment Anwendungen spielt das Mobiltelefon eine zentrale Rolle. Die Idee dahinter: Das Handy hat jeder immer und überall dabei. Will ein Käufer eine Zahlung tätigen, muss er nach Erhalt einer SMS oder eines eingehenden Anrufs diese Zahlung durch Eingabe einer PIN bestätigen. Zu den bekanntesten und etabliertesten Anbieter dieses Zahlungsverfahrens gehören Street Cash [27] oder paybox [8]. paybox erfreut sich v. a. in Österreich einer grossen Beliebtheit, und es können auch eine Vielzahl von Real World Angeboten damit bezahlt werden. So zum Beispiel Taxifahrten, Tickets für den öffentlichen Verkehr oder sogar Konzertkarten.

National konnten sich solche mPayment Systeme bereits etablieren. Eine Zahlung per mPayment zu tätigen ist sehr einfach, und das Verfahren ist auch ausreichend sicher. Spätestens wenn sich die verschiedenen Mobilnetzbetreiber auf ein gemeinsames System einigen können, steht deshalb dem Erfolg von mPayment nichts mehr im Weg.

Verrechnung über monatliche Abrechnung

Eine andere Idee ist die Bezahlung von Inhalten über eine monatliche Abrechnung. Die Firma Firstgate konnte nach Deutschland mittlerweile in verschiedene anderen, v. a. europäischen Ländern eine grosse Zahl von Internet-Inhalteanbietern gewinnen, welche eine Bezahlung mit click&buy [23] akzeptieren. Firstgate arbeitet dabei teilweise mit internationalen Partner zusammen, beispielsweise mit der Swisscom oder British Telecom. Möchte ein Kunde bestimmte kostenpflichtige Inhalte ansehen, muss er sich mit seinem Benutzernamen und Kennwort einloggen. Seinem Konto wird danach der entsprechende Betrag belastet.

Die grosse Attraktivität von click&buy besteht im breiten und interessanten Angebot verschiedener Anbieter.

10.3.6 Fazit

Der Bezahlvorgang ist ein wichtiger Bestandteil der Abwicklung von Geschäften im Internet. Entsprechend sind geeignete Zahlungsmethoden von entscheidender Bedeutung für den Erfolg von eBusiness. Zu den wichtigsten grundsätzlichen Anforderungen an eine Zahlungsmethode gehören Sicherheit, Bedienerfreundlichkeit oder Kosten (Spesen) pro Transaktion. Zur Lösung dieser Probleme wurden verschiedene Ansätze vorgeschlagen und daraus verschiedene Zahlungsverfahren entwickelt, welche den oben genannten Anforderungen genügten. Wie sich gezeigt hat, genügt es jedoch nicht nur diese technischen Ansprüchen zu erfüllen. Von eben so grosser Bedeutung ist es, dass eine Zahlungsmethode zusätzlich breit akzeptiert ist, das heisst einerseits von vielen Käufern benutzt wird, und dass die Bezahlung mit dieser Methode bei einer grossen Anzahl von Verkäufern möglich ist. Da das Internet ein globales Medium ist, sollen die Zahlungsverfahren zudem ebenso global einsetzbar sein.

Im Macropayment-Bereich haben sich in der Zwischenzeit einzelne Zahlungsmethoden wie z. B. Kreditkarte, Rechnung oder Nachnahme etabliert und bewährt. In diesem Bereich ist deshalb nicht mit radikalen Reformen und wesentlich veränderten Zahlungsmethoden zu rechnen.

Im Micropayment-Bereich konnte sich im Gegensatz dazu noch keine der vorgestellten Zahlungsmethoden entscheidend durchsetzen, obwohl einige viel versprechende und interessante Verrechnungsysteme entwickelt wurden. Das Hauptproblem bei kleinen Zahlungsbeträge im Micropayment-Bereich besteht darin, dass die Kosten, die eine Transaktion verursacht im Verhältnis zum Zahlungsbetrag nicht zu gross werden darf. Um dies zu erreichen, werden fast bei allen vorgeschlagenen Methoden die einzelnen Zahlungen kumuliert und zusammen beglichen. Es wird sich nicht zwingend das technisch am meisten ausgereifte Zahlungsverfahren durchsetzen, sondern dasjenige, welches eine genügend hohe Verbreitung bei Käufern und Verkäufern findet. Am meisten Chancen haben dabei die Systeme, die von Organisationen mit gutem Ruf und bereits grossem Kundenstamm getragen beziehungsweise mitgetragen werden (zu diesen Organisationen könnten beispielsweise Telekom-Anbieter oder Internetfirmen wie Ebay und Google gehören).

Kann sich ein Verfahren durchsetzen, könnte das zu neue Geschäftsmodellen für die Anbieter von Internet-Inhalten führen. Viele dieser Anbieter haben bis heute ihr Angebot entweder gratis zur Verfügung gestellt, oder haben sich über Werbung finanziert, weil es keine geeigneten Micropayment-Zahlungsmethoden gab. Das könnte schliesslich wiederum dazu führen, dass mit der Zeit die Mentalität verschwinden könnte, dass im Internet alles gratis erhältlich ist.

10.3.7 Ausblick

Letzten Endes wird es der Kunde sein, der entscheidet welches System sich wann durchzusetzen vermag. Dabei sind wie in dieser Arbeit mehrfach erwähnt die zentralen Punkte die Sicherheit und Einfachheit eines Systems. Oftmals existieren ausgereifte Varianten, die sich aber bei der technisch nicht sonderlich interessierten Masse nicht durchzusetzen vermögen. Das perfekte System könnte demnach durch seine, von den Endbenutzern anerkannte Sicherheit und absolut intuitiver Bedienung eine neue Ära im Internet einleiten.

Bibliography

[1] Website der deutschen GeldKarte, [online] 2005, http://www.geldkarte.de (Link besucht am 24.01.2005)

[2] Online Shop Reiner SCT, [online] 2005, https://www.chipkartenleser-shop.de/shop/rsct.html (Link besucht am 24.01.2005)

[3] Website der österreichischern billiteasy, [online] 2005, http://www.billiteasy.com (Link besucht am 24.01.2005)

[4] Website von Click and buy Schweiz, [online] 2005, http://www.clickandbuy.ch (Link besucht am 24.01.2005)

[5] Website Firstgate, [online] 2005, http://www.firstgate.com (Link besucht am 24.01.2005)

[6] Website Ebay, [online] 2005, http://www.ebay.ch (Link besucht am 24.01.2005)

[7] Schweizer Website von PayPal, [online] 2005, http://www.paypal.com/ch (Link besucht am 24.01.2005)

[8] Website der paybox.net AG, [online] 2005, http://www.paybox.net (Link besucht am 24.01.2005)

[9] verschiedene Diskussionen über unlautere Geschäftspraktiken der Firma Moxmo, [online] 2005, http://forum.computerbetrug.de/viewtopic.php?p=53685 oder http://kopfkrebs.de/Moxmo_Deutschland_AG (Link besucht am 24.01.2005)

[10] deutsche Informationsseite, [online] 2005, http://www.dialerschutz.de/premium-sms.php , 02.02.05

[11] Deutsche Bank Research: "'ePayments: zeitgemäße Ergänzung traditioneller Zahlungssysteme"', [online] 2005, http://www.ecin.de/zahlungssysteme/onlinepayment2/ (Link besucht am 24.01.2005)

[12] Robben Mathias: "'ePayment: Alte Besen kehren noch am besten"', [online] 2005, http://www.ecin.de/zahlungssysteme/epayment/(Link besucht am 24.01.2005)

[13] Krempl Stefan, "'eCash und Co: Das waren Kopfgeburten"', [online] 2005, http://www.heise.de/tp/r4/artikel/7/7477/1.html (Link besucht am 24.01.2005)

[14] Website von Swisscom Fixnet Easyp@y, [online] 2005, http://www.easypay.ch/ (Link besucht am 24.01.2005)

[15] Burns, Joe Ph.D.: What is PayPal, [online] 2005, http://www.htmlgoodies.com/beyond/paypal.html (Link besucht am 24.01.2005)

[16] Robben, Matthias: Online Payment: Bleibt alles beim Alten?; ECIN - Electronic Commerce Info Net, Dortmund, [online] 2005, http://www.ecin.de/zahlungssysteme/onlinepayment/index-3.html (Link besucht am 24.01.2005)

[17] Robben, Matthias: Paid Content: Wer soll das bezahlen...?; ECIN - Electronic Commerce Info Net, Dortmund, [online] 2005, http://www.ecin.de/state-of-the-art/paidcontent/ (Link besucht am 24.01.2005)

[18] Heng, Dr. Stephan: E-Payment-Systeme: Frische Brise im E-Business; Deutsche Bank Research; Düsseldorf, 06.10.2004, [online] 2005, http://www.dbresearch.de/PROD/DBR_INTERNET_DE-PROD/PROD0000000000180079.pdf (Link besucht am 23.01.2005)

[19] Website von MicroMoney, [online] 2005, http://www.micromoney.de/ (Link besucht am 24.01.2005)

[20] Website von PayNet, [online] 2005, http://www.paynet.ch/ (Link besucht am 24.01.2005)

[21] Website von PayPal international, [online] 2005, http://www.paypal.com/ch/cgi-bin/webscr?cmd=p/gen/about-outside (Link besucht am 24.01.2005)

[22] Website von paysafecard, [online] 2005, http://www.paysafecard.com/ (Link besucht am 24.01.2005)

[23] Webseiten von Swisscom click&buy und Firstgate, [online] 2005, http://www.clickandbuy.ch/ (Link besucht am 24.01.2005) http://www.firstgate.com/ (Link besucht am 24.01.2005)

[24] Spann, Martin; Pfaff, Donovan: Electronic Bill Presentment and Payment (EBPP), [online] 2005, http://www.ecommerce.wiwi.uni-frankfurt.de/skiera/publications/ebpp_spann_pfaff.pdf (Link besucht am 23.01.2005)

[25] Reichmayr, Christian: ePayment kooperative Zahlungsprozesse und eServices im Internet; Institut für Wirtschaftsinformatik, Universität St. Gallen, 14.05.2001

[26] Website von Postfinance yellowbill, [online] 2005, http://www.yellowbill.ch (Link besucht am 24.01.2005)

[27] Website von Street Cash, [online] 2005, http://www.streetcash.de/ (Link besucht am 24.01.2005)

Chapter 11

Migration to IPv6

Andreas Drifte, Silvan Hollenstein, Roger Grütter

Das IPv4 leistete in den vergangenen Jahren gute Dienste. Da aber der Adressraum zur Neige geht, beschloss die IETF ein neues IP-Protokoll entwickeln zu lassen, das die Adressen mit 128 Bit statt 32 Bit darstellt. IPv6 hat viele Merkmale von IPv4 übernommen, aber es unterscheidet sich in Details erheblich. Da eine Ad-Hoc Umstellung von IPv4 zu IPv6 wegen der Komplexität der Netzinfrastukur von heute auf morgen nicht möglich ist, wurden Migrationsstrategien entwickelt, um einen fliessenden Übergang auf IPv6 zu vollziehen und gleichzeitig Kompatibilität zu IPv4 zu gewährleisten. Schrittweise werden IPv6-Inseln erstellt welche mittels Dual-Stack-Routern, welche beide Protokolle beherrschen, ans IPv4-Netzwerk angebunden werden. Mittels Tunnelstrategien wird der Transport von IPv6-Paketen über das bestehende IPv4-Netzwerk ermöglicht. Es stellt sich die Frage ob überhaupt zwingend ein neues Protokoll benötigt wird, da die meisten Vorteile von IPv6 auch bei IPv4 auf höherer Ebene implementiert werden konnte. Wegen des knappen Adressraums von IPv4, welche besonders für Asien nicht genügend IPv4-Adressen bereitstellen kann, ist ein Umstieg zwangsläufig notwendig. Ob alle ISP's und Firmen auf den IPv6-Zug aufspringen werden ist fraglich, da die Umstellung enorme Kosten verursacht. Sowohl die Router wie auch alle Computer müssen auf IPv6 umgestellt werden, ohne dass daraus ein kurzfristiger Nutzen erzielt wird. Bisher investierten nur wenige ISP und Unternehmen in IPv6. In der Schweiz sind das bespielsweise Dolphins, Cyberlink, Solnet und die Switch. Längerfristig, vor allem wenn mehrere Haushaltsgeräte eine IP-Adresse benötigen, wird IPv6 das Protokoll der Zukunft.

Inhaltsangabe

11.1 Einleitung

Für den Transport von Daten wird nun seit über 25 Jahren das Internet Protokoll in der Version 4 (IPv4) eingesetzt. Zu diesem Zeitpunkt war dieses Protokoll für einen relativ kleinen Kreis von Nutzern, wie z.B. Hochschulen, Regierungen oder das Militär, vorgesehen. Zu Beginn der 90er Jahre hat das Internet rasch zu wachsen begonnen und wächst auch heute noch von Jahr zu Jahr rasant weiter. Die Ursprünglichen Anforderungen an das Internet Protokoll haben sich durch dieses rasante Wachstum, aber auch durch bestimmte Designfehler verändert. Zudem stellen neue multimediale Applikationen wie Video-on-demand oder Internettelefonie neue, veränderte Anforderungen an das Internetprotokoll. Diesen neuen Anforderungen ist das IPv4 nicht mehr gewachsen und es musste in absehbarer Zukunft eine Lösung dafür gefunden werden, wie das Internet Protokoll den neuen Anforderungen gerecht werden kann. Das Resultat aus diesen Bestrebungen ist das Internet Protokoll in der Version 6.

Doch nur mit der Enwicklung eines neuen Protokolls ist es nicht getan, es müssen auch Strategien ausgearbeitet werden, die eine Migration von IPv4 zu IPv6 möglich machen. Durch die weltweite Verbreitung und die immense Anzahl an Geräten, die bei einer Umstellung neu konfiguriert oder sogar updated werden müssen, müssen Strategien gefunden werden, die es ermöglichen beide Protokollversionen parallel betreiben zu können. Durch solche geeignete Strategien ist es möglich, dass beide Protokollversionen miteinander existieren können und auch gegenseitig Daten austauschen können. Somit ist es möglich Schritt für Schritt einzelne Teilnetze zu IPv6 zu migrieren.

In unserer Arbeit werden wir zuerst kurz das IPv4 vorstellen und dabei die entstandenen Probleme mit diesem Protokoll in der heutigen Zeit aufzeigen. Danach soll das neue Internet Protokoll in der version 6 vorgestellt werden. Abschliessend beschäftigen wir uns mit den verschiedenen Migrationsstrategien, die es ermöglichen das IPv4 Schritt für Schritt durch das IPv6 abzulösen. Zum Schluss möchten wir noch einige wirtschafliche Aspekte beleuchten, wie etwa die Kosten oder die Dauer einer weltweiten Migration.

11.2 Das Internet Protokoll Version 4

Es existieren heute viele verschiedene Netzwerke, die alle auf unterschiedlichen Technologien beruhen. Beispiele dafür sind Ethernet-, Token-Ring- oder FDDI-Netzwerke. Die Pakete (*Frames*), die in diesen Netzwerken verschickt werden, weisen kein einheitliches Format auf. Ein globales Internet ist deshalb nur möglich, wenn die lokalen Unterschiede verborgen bleiben, und genau dies ist die Aufgabe eines Internet Protokolls.

Mit der Einführung der Protokollfamilie TCP/IP begann der Siegeszug von vernetzten Systemen. Das aktuelle Internet Protokoll Version 4 (IPv4) hat lange Zeit für ein erfolgreiches Internetworking gesorgt, und tut es auch heute noch. In dieser Arbeit wird deshalb als Einführung noch das IPv4 vorgestellt, und soll so auch dazu beitragen, die beiden Internet Protokolle, IPv4 und IPv6, miteinander vergleichen zu können.

11.2.1 Das IPv4-Adressformat

In TCP/IP wird ein beliebiges Computersystem, das mit dem Internet verbunden ist, als *Host* bezeichnet. Die Hosts sind über Router miteinander verbunden. Die Router müssen also auch mit TCP/IP-Protokollsoftware ausgestattet sein. Die Router müssen aber nicht alle Schichten des TCP/IP-Protokolles implementieren, vor allem nicht die Applikations-Schicht, weil auf Routern keine Anwendungen, wie zum Beispiel Dateitransfer, laufen. Im Normalfall hat ein Host eine IP-Adresse. Es können aber auch mehr sein, falls der Host an mehrere Netzwerke angeschlossen ist. Ein Router hat immer mindestens zwei IP-Adressen: Für jedes Netzwerk, an das er angeschlossen ist, erhält er eine IP-Adresse [1].

Die Aufgabe von IP ist die Übermittlung von Paketen (*Datagrammen*) von einem Host an einen anderen. IP übernimmt aber keine Garantie, dass ein Paket auch wirklich beim Empfänger ankommt. Für eine zuverlässige Übertragung sind Protokolle der höheren Schichten wie zum Beispiel TCP zuständig. Die Adressierung erfolgt über 32 Bit lange Binärzahlen. Da dies nicht unbedingt gut lesbar ist, wird die Adresse in 4 mal 8 Bits unterteilt, wobei ein einzelner 8-Bit-Block durch eine Zahl zwischen 0 und 255 dargestellt wird, und die einzelnen Blöcke durch einen Punkt getrennt werden. Eine IPv4-Adresse könnte also folgendermassen aussehen: 231.120.0.1

Adressklassen

IP-Adressen sind hierarchisch gegliedert: Sie setzen sich immer aus einem Präfix und einem Suffix zusammen. Das Präfix identifiziert ein Netzwerk und das Suffix einen bestimmten Host in diesem Netzwerk. Dies hat den Vorteil, dass nur die Präfixe in die Routing-Tabellen eingetragen werden müssen. Früher wurden die IP-Adressen in fünf Klassen unterteilt (*Classful IP Addressing*). Wie die Grafik 11.1 zeigt, besteht das Präfix bei der Klasse A aus der ersten Zahl, bei der Klasse B aus den ersten zwei Zahlen und bei Klasse C aus den ersten drei Zahlen. Klasse D ist für Multicast Gruppen, und Klasse E für zukünftige Anwendungen reserviert.

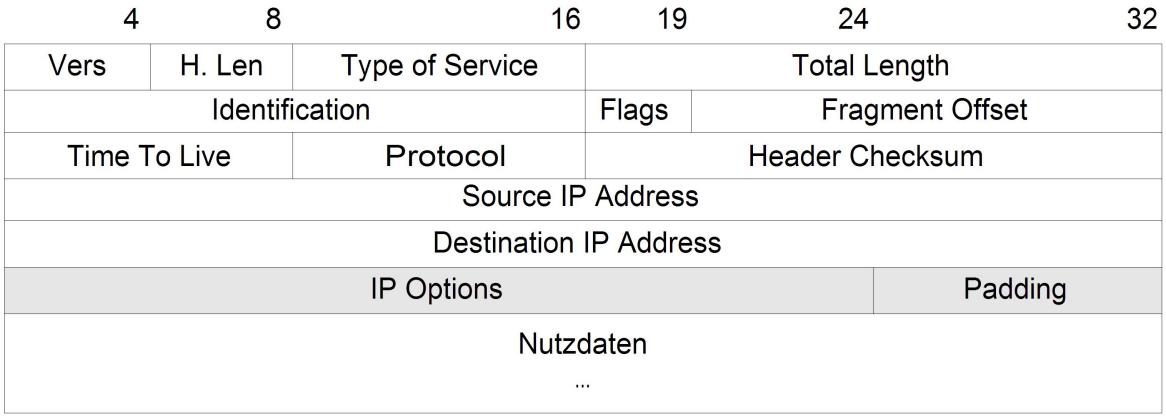

Figure 11.1: Die fünf Adressklassen

Die Aufteilung in verschiedene Klassen hatte den Vorteil, dass für unterschiedlich grosse Netzwerke die jeweils passende Klasse gewählt wurde. Mit dem enormen Wachstum des Internet ergaben sich aber bald Probleme. Für die meisten mittelgrossen Unternehmen sind Klasse B-Netze zu gross, und Klasse C-Netze mit maximal 254 Hosts zu klein [2]. Auf der anderen Seite gab es viele B-Netze mit weniger als 65'534 Hosts, und somit ungenutzten Adressen.

Subnetz- und klassenlose Adressierung

Durch das Classful IP Addressing zeichnete sich bald eine Adressknappheit ab. Man kam auf die Idee [2], die Aufteilung in Präfix und Suffix nicht nur an den Byte-Grenzen, sondern an beliebiger Stelle zu vollziehen. Die Schreibweise für eine IP-Adresse sieht dann wie folgt aus: 192.168.0.0/16, was in diesem Fall einer Klasse B-Adresse entsprechen würde. Aber auch folgende Adresse ist zulässig: 192.168.64.0/23 steht für die Adressen von 192.168.64.0 bis 192.168.65.255. Die Schreibweise mit Schrägstrich, *CIDR* genannt (Classless Inter-Domain Routing), wird im Computer anders gehandhabt. Zu jeder IP-Adresse existiert eine 32-Bit-Maske, die angibt, wieviele Bits lang das Präfix ist. Die Maske für /23 wäre beispielsweise 255.255.254.0. Jedes Bit, das noch zum Präfix gehört, wird als 1 notiert, der Rest als 0. Der Grund für eine Bit-Maske ist der, dass Computer aus der Maske und der Adresse mit nur zwei Maschineninstruktionen das Präfix berechnen können [1].

In den Routing-Tabellen wird zusammen mit der IP-Adresse immer auch die zugehörige Maske gespeichert. Kommt nun ein IP-Paket mit einer bestimmten Adresse bei einem Router an, wird es in einer logischen UND-Operation mit jeder Maske verknüpft. Bei dieser Operation wird das Präfix aus der IP-Adresse extrahiert. Stimmt das Präfix dann mit der Adresse neben der vorher verwendeten Maske, wird das Paket an den nächsten Hop gesendet.

Zusätzlich wird die Tabelle so aufgebaut, dass automatisch die Übereinstimmung mit dem längsten Präfix (Longest Prefix Match) genommen wird. Nehmen wir an, dass es einen Eintrag für ein Netz mit der Adresse 192.168.0.0/16 gibt, siehe Tabelle 11.1. Es kann nun vorkommen, dass es für ein bestimmtes Subnetz eine andere Route gibt, z.B. für 192.168.24.0/24. Kommt nun ein Paket mit der Zieladresse 192.168.24.2 an, würde es bereits mit dem ersten Eintrag eine Übereinstimmung finden und an 128.14.0.9 weitergeleiten werden. Eine Routing-Tabelle ist deshalb so geordnet, dass die längeren Präfixe zuerst erscheinen, und somit automatisch die Bedingung des Longest Prefix Match erfüllt ist.

Spezielle IPv4 Adressen

Neben den für Hosts reservierte Adressen gibt es noch eine Reihe von speziellen Adressen, die zum Beispiel für ein Netzwerk oder Computergruppen gelten [1].

Netzwerkadresse Nach dem IP-Standard steht die Host-Adresse 0 für das Netzwerk.
"'Die Netzwerkadresse bezieht sich auf das Netzwerk selbst und nicht auf die daran

Table 11.1: Verwendung von Masken in Routern, Tabelle aus [1]

Ziel	**Maske**	**Nächster Hop**
192.168.24.0	255.255.255.0	128.14.0.9
192.4.10.0	255.255.255.0	130.111.0.1
192.168.0.0	255.255.0.0	128.15.14.2
128.1.0.0	255.255.0.0	Direkte Zustellung
30.0.0.0	255.0.0.0	40.0.0.7

angeschlossenen Host-Rechner. Deshalb sollte die Netzwerkadresse nie als Zieladresse in einem Paket erscheinen."' [1]

Beispiel: 192.168.0.0/16

Gerichtete Broadcast-Adressen Ein Suffix, das nur aus Einsen besteht, bedeutet ein gerichtetes Broadcast. Dies kann nützlich sein, wenn ein Paket an alle Hosts gesendet werden soll, die an einem bestimmten Netzwerk angeschlossen sind.

Beispiel: 192.168.255.255/16

Begrenzte Broadcast-Adressen Diese Adresse besteht vollständig nur aus Einsen, also 255.255.255.255. Sie wird von einem Computer beim Systemstart benutzt, wenn ihm die Netzwerkkennung noch nicht bekannt ist.

This Computer Diese Adresse wird vom Computer beim Systemstart benutzt, wenn er noch keine IP-Adresse zugewiesen bekommen hat. Sie besteht nur aus Nullen, also 0.0.0.0.

Schleifenadressen Diese Adressen sind für Testzwecke vorgesehen. Falls zwei Anwendungsprogramme, die üblicherweise über ein Netzwerk miteinander kommunizieren, zum Test auf nur einem Computer ausgeführt werden sollen, brauchen sie trotzdem IP-Adressen. Dafür sind alle Adressen mit dem Präfix 127/8 reserviert.

Private Adressen [5] Bestimmte Adressbereiche sind nur für private Zwecke bestimmt und werden daher nie im Internet geroutet. Sie bieten sich vor allem an bei Firmen, die auf Grund der Adressknappheit nicht mehr genügend IP-Adressen erhalten, oder aber aus Sicherheitsgründen. Auf die privaten Adressen wird noch im Abschnitt 'Routing in IPv4 - NAT' näher eingegangen. Die privaten Adressbereiche sind:

10/8	10.0.0.0	bis 10.255.255.255
172.16/12	172.16.0.0	bis 172.31.255.255
192.168/16	192.168.0.0	bis 192.168.255.255

11.2.2 Der IPv4-Header

Der IP Header ist dafür verantwortlich, dass das IP Paket im Internet richtig geroutet wird. Die wichtigsten Funktion sind dabei die Ziel- und Quelladresse. Aber auch weitere

Steuerelemente sind im Header enthalten. Im folgenden wird die Abbildung 11.2 und die einzelnen Steuerelemente des IPv4 Headers erklärt. Die Angaben dazu sind aus [1] und [4].

Vers: Hier wird die Version des Internet Protokolls eingetragen. Momentan steht hier also eine 4.

Header Length: Die Header Länge wird in 32-Bit Blöcken gezählt, und beträgt somit ohne IP Options 5.

Type of Service: Dieses Feld enthält die Angabe, ob das Datagramm über eine Route mit minimaler Verzögerung, oder eine mit maximalem Durchsatz gesendet werden soll.

Total Length: 16 Bit beschreiben die Länge des gesamten IP Pakets, also inklusive Header und Nutzdaten. Die Gesamtlänge kann demnach höchstens 65'535 Bytes betragen.

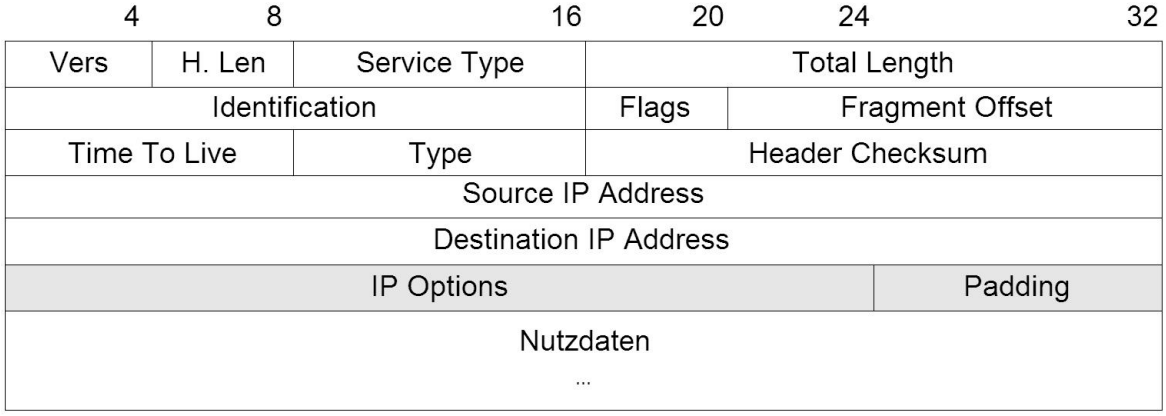

Figure 11.2: Der IPv4 Header

Identification: Jedes IP-Paket erhält eine 16 Bit lange Identifikationsnummer. Wenn ein Router ein IP-Paket fragmentieren muss, setzt er in alle neu generierten Header der Fragmente eine Kopie der Identification. Der Empfänger kann so die Fragmente richtig zuordnen.

Flags: Das Feld Flags kennzeichnet, ob ein Datagramm ein Fragment oder ein vollständiges Datagramm ist.

Fragment Offset: Falls ein IP-Paket nur ein Fragment ist, wird hier angezeigt, an welche Stelle des Datagramms das Fragment gehört.

Time To Live: Dieses Feld enthält eine Zahl zwischen 1 und 255 und wird beim Generieren des IP-Pakets gewählt. Jeder Router, der das Paket weiterleitet, senkt diesen Wert um 1. Ist der Wert bei 0 angelangt, verwirft der Router das Paket und sendet eine ICMP-Nachricht an den Absender zurück. Somit wird garantiert, dass das Paket nicht in eine Endlosschlaufe geraten kann.

Protocol: In diesem Feld steht der Datentyp, den das IP-Paket transportiert, zum Beispiel TCP oder UDP.

Header Checksum: Dieses Feld ist für die Datenintegrität des Headers zuständig. Der gesamte Header (ausser Header Checksum) wird in 16 Bit-Wörter aufgeteilt. Dann werden alle Wörter zusammengezählt, und falls es einen Überlauf gibt, also das 17. Bit eine 1 ist, wird diese 1 hinzugezählt. Am Schluss wird von dieser Addition das Komplement genommen. Der Empfänger des Pakets führt dieselbe Addition durch, zählt aber die Checksum noch hinzu, und erhält dann als Ergebnis null, falls das Paket nicht verändert wurde. Jeder Router muss diesen Wert wieder neu setzen, weil einige Felder sich ändern können, z.B. Time To Live.

Source / Destination IP Address: Source IP Address enthält die Adresse des Absenders, Destination IP Address die Adresse des Ziels.

IP Options / Padding: Hier können verschiedene zusätzliche Optionen festgelegt werden. Das Feld Padding wird benutzt, um auf 32 Bit aufzufüllen. Wenn keine Optionen eingefügt werden, endet hier der IP-Header.

11.2.3 Routing in IPv4

Bevor auf das Routing von IP-Paketen eingegangen wird, soll noch erklärt werden, wie das Auffinden eines Hosts innerhalb eines LANs geschieht, denn in einem LAN werden die Hosts nicht mit IP-Adressen angesprochen, sondern mit ihrer Hardwareadresse. Und um die IP- auf die Hardwareadresse abzubilden, gibt es das Address Resolution Protocol (ARP).

ARP - Bindung von Protokolladressen

Wenn ein Computer einem anderen über LAN eine Nachricht schicken möchte, und dabei nur die IP-Adresse des anderen kennt, muss er mit einer ARP-Nachricht zuerst die physische Ethernet-Adresse herausfinden. Er sendet dazu einen ARP-Request-Broadcast an alle Systeme, die an das Ethernet-Segment angeschlossen sind. Diese Nachricht enthält die gesuchte IP-Adresse. Das System mit der gesuchten IP-Adresse sendet dann eine ARP-Reply-Nachricht mit der Hardware-Adresse (z.B. MAC-Adresse) an den Absender zurück. [3]

Diese Nachricht hat zwar eine genaue Spezifikation, ist aber trotzdem sehr flexibel. Die ARP-Nachricht enthält in den ersten zwei Feldern Identifikatoren für den Hardwaretyp und die Protokolladresse. Danach folgen zwei 8-Bit-Felder, die die Länge der Hardware- und Protokolladressen anzeigen [1]. Eine IP-Adresse könnte demnach bis zu 255 Bit lang sein.

Statisches Routing

Statisches Routing trifft man vor, wenn ein Router fixe Einträge in seiner Routing Tabelle hat. Ein Administrator muss diese Einträge zu Beginn oder bei Veränderungen manuell eingeben. Die Vorteile von statischem Routing sind: Es braucht keine Routing-Software, es wird keine Bandbreite beansprucht, und die CPU wird nicht benötigt. Dafür gibt es aber entscheidende Nachteile in der Flexibilität: Netzwerkstörungen und Topologieänderungen können nicht berücksichtigt werden [1]. Für die Router in einem unüberschaubaren Netzwerk kommt das statische Routing nicht in Frage, weil niemand in der Lage wäre, solche Routing Tabellen zu pflegen. Statisches Routing wird vor allem bei den Hosts verwendet, wo meistens zwei Einträge in der Tabelle genügen, ein Eintrag für das lokale Netzwerk, und ein Default-Wert, der auf den Standard-Gateway verweist. Auf Firewall-Komponenten werden meistens auch statische Tabellen verwendet, da dynamische Protokolle Angriffspunkte darstellen. [3]

Dynamisches Routing

Beim dynamischen Routing tauschen die Router untereinander Informationen aus, die es ihnen ermöglicht, für jede beliebige IP-Adresse den nächsten Hop zu bestimmen. Der Router aktualisiert dabei fortlaufend seine Routing-Tabelle. Auf das gesamte Internet lässt sich das aber nicht anwenden, denn wenn alle Router Informationen miteinander austauschen würden, wäre das Netzwerk schnell überlastet. Deshalb gibt es die zweischichtige Routing-Hierarchie: Einzelne Gruppen von Routern werden zu autonomen Systemen (AS) zusammengefasst. Innerhalb eines AS wird ein Interior Gateway Protocol verwendet. Mindestens ein Router dieses AS fasst die gesamten Informationen des Netzwerks zusammen und tauscht diese über ein Exterior Gateway Protocol mit anderen AS aus.

Das Routing innerhalb eines AS wird auch als Intra Domain Routing, und das Routing zwischen AS als Inter Domain Routing bezeichnet. Beide Verfahren werden näher im Abschnitt 3.3 über Routing in IPv6 beschrieben.

Network Address Translation

Im Abschnitt über spezielle IPv4-Adressen wurde bereits erwähnt, dass bestimmte Adressbereiche für die private Nutzung reserviert sind, und auch niemals im Internet geroutet werden. Man darf also im eigenen Netzwerk frei über diese nach RFC 1918 [5] spezifizierten Adressen verfügen. Damit vom internen Netz Pakete auch ins Internet gesendet werden können, braucht es eine Firewall oder einen Router, der die internen in externe Adressen und umgekehrt übersetzt. Man nennt dies die Network Address Translation (NAT).

Die folgende Figur aus [3] zeigt ein privates Netzwerk mit der Adresse 192.168.0.0/16, das über ein Firewall-Gateway mit dem Internet verbunden ist. Das Gateway hat die IP-Adresse 194.120.66.1 vom ISP zugewiesen bekommen. Es verwaltet eine Tabelle mit

Zuordnungen von IP-Adresse und Portnummer zu neuer IP-Adresse und neuer Portnummer. Wenn zum Beispiel ein Paket vom Rechner mit der internen Adresse 192.168.2.1 beim Gateway ankommt, schreibt es den IP-Header um, und setzt als Absender die offizielle, vom ISP erhaltene Adresse, 194.120.66.1 und Portnummer 33100. Für jeden Rechner gibt es dann eine Menge von Portnummern, die eindeutig zugewiesen werden können.

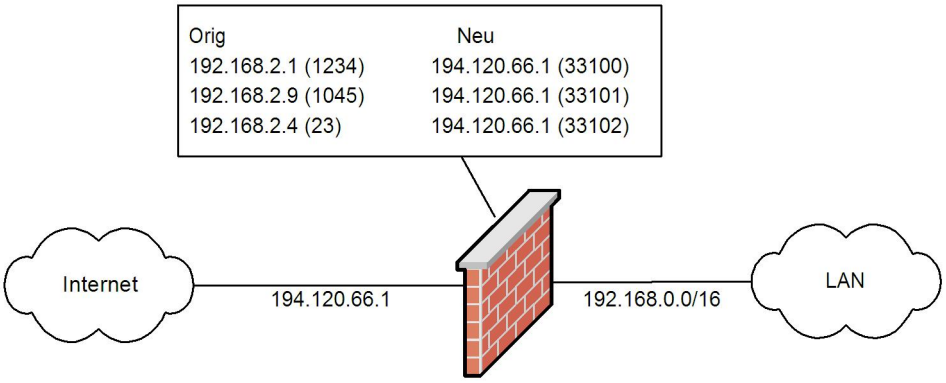

Figure 11.3: Network Address Translation

Weil NAT mit Portnummern arbeitet, funktioniert diese Technik nicht mit IP-Paketen, die in den Nutzdaten keine Portnummer enthalten. Vielmehr braucht es ein höheres Protokoll wie TCP oder UDP.

Mit NAT hat man nun die Möglichkeit, mehrere Rechner an eine einzige IP-Adresse zu binden. Dies hat den Vorteil, dass im gesamten Internet weniger IP-Adressen zur Verfügung stehen müssen als Hosts, und entschärft dadurch das Problem der Adressknappheit, ganz beseitigen kann es das aber nicht. RFC 1918 [5] von 1996 spricht denn auch zum ersten die Adressknappheit an, und zum zweiten das Problem, dass die Router immer mehr Overhead befördern müssen.

11.2.4 Warum braucht es ein neues Internetprotokoll?

Probleme mit IPv4

Wie bereits mehrfach erwähnt, ist das grösste Problem bei IPv4 die Adressknappheit. Gegen Ende 2002 waren über 70 Prozent der IP-Adressen verteilt, wovon 74 Prozent nach Nordamerika, 17 Prozent nach Europa und lediglich 9 Prozent nach Asien gingen. [6] Vor allem in Asien ist momentan der Druck nach mehr IP-Adressen so gross, dass allem Anschein nach IPv6 sich dort am schnellsten etablieren wird.

Mit NAT konnte die Adressknappheit zwar entschärft werden, hat aber neue Probleme mit sich gebracht. NAT ist uni-direktional, ein Host hinter einem Gateway kann zwar das Internet erreichen, ist aber selber unsichtbar. Das heisst, bei allen Anwendungen, bei denen der Host direkt angesprochen werden soll, bräuchte er eine eindeutige und fixe Adresse. Solche Anwendungen könnten zum Beispiel IP-Telefonie, Videophone, Computerspiele, oder jegliche interaktive Anwendungen sein.

Ein weiteres Problem, das momentan besteht, ist die Überlastung der Router. Auf den Backbone-Routern sind zum Teil Routing-Tabellen mit bis zu 80'000 Einträgen vorhanden. Ein ankommendes Paket muss also im Durchschnitt mit 40'000 Einträgen verglichen werden, bevor es geroutet werden kann. Die Ursache für dieses Problem liegt hauptsächlich in der fehlenden topologischen Struktur des Internets. Nehmen wir ein Beispiel:

- Sei das Adresspräfix 107.211.0.0/17 im Land A,
- und das Adresspräfix 107.211.128.0/17 im Land B

Ein Router im Land C muss also zwei Einträge haben für sehr ähnliche Präfixe. Wünschenswert wäre natürlich, dass alle Adressen mit dem Präfix 107.211.0.0/16 in einem Land wären, so dass die Einträge auf einen Bruchteil reduziert werden können. Um es allgemein auszudrücken: Die ersten x Bits einer Adressen müssen auf eine geographische Region verweisen (z.B. ein Land), die nächsten y Bits kann dieses Land wieder auf Regionen (oder Provider) aufteilen und so fort. Wie im nächsten Abschnitt gezeigt wird, ist dies eben möglich mit IPv6.

Anforderungen an ein neues Internetprotokoll

Die zwei wichtigsten Anforderungen an ein neues Protokoll sind aus den bereits genannten Gründen ein grösserer Adressraum und schnelleres Routing. Daneben gibt es aber auch noch weitere Bedürfnisse der Benutzer, die mit dem jetzigen Internet Protokoll nicht zu erfüllen wären. So ist es momentan nicht möglich, Multicast richtig einzusetzen, ausser man befindet sich im selben Subnetz. In Zukunft wird es aber vielleicht vermehrt Broadcast, bzw. Multicast über das Internet geben. Eine Möglichkeit wäre eine Art Pay-per-View, wo also zu einem bestimmten Zeitpunkt der neuste Kinofilm über das Netz wandert, und nur ausgewählten Benutzern (z.B. zahlenden Kunden) zugesandt wird. Anstatt dass sich jeder unabhängig einen Film ansieht, könnte mit Multicast Bandbreite gespart werden.

In die gleiche Richtung geht der Wunsch nach Quality of Service (QoS). So soll bei Voice over IP möglichst kurze Verzögerung , und beim Betrachten eines Filmes hohe Bandbreite sichergestellt werden.

Eine weitere Anforderung ist der Wunsch nach Mobilität. So soll in Zukunft jedes mobile Telefon seine eigene IP-Adresse besitzen, und hierbei werden hohe Anforderungen an das Protokoll gestellt, da es Techniken wie Roaming und Handover beherrschen muss. Aber auch andere Geräte wie Kühlschrank, Waschmaschine oder das Navigationsgerät im Auto sollen eines Tages ihre eigene Internetadresse haben, was natürlich deutlich macht, dass jeder Mensch mehrere IP-Adressen besitzen muss.

11.3 Das Internet Protokoll Version 6

Im letzten Abschnitt haben wir gesehen, dass ein neues Internetprotokoll, zumindest in der langfristigen Sicht, benötigt wird. Die Notwendigkeit eines neuen Internetprotokolls

wurde auch von der IETF (Internet Engineering Task Force) erkannt und deshalb forderte diese im Dezember mit dem RFC 1550 [7] die Internetgemeinde dazu auf, Vorschläge für ein neues Internetprotokoll zu machen. Aufgrund dieser Anfrage wurde eine Vielzahl von Vorschlägen eingereicht. Diese Vorschläge reichten von geringen Veränderungen beim bestehenden IPv4 [4] bis zur vollständigen Ablösung des IPv4 durch ein neues Protokoll. Aus den eingereichten Vorschlägen wurde von der IETF das Simple Internet Protocol Plus (SIPP) ausgewählt, das sich aus den Vorschlägen von Deering [8] und Francis [9] zusammensetzt. Die offizielle Versionsnummer des neuen Internetprotokolls ist sechs. Der Sprung von zwei Versionsnummern ist nicht etwa darauf zurück zu führen, dass sich das neue Protokoll total vom alten unterscheidet, sondern dass die Protokollnummer 5 schon für ein experimentelles Protokoll verwendet wurde. Die folgenden Beschreibungen von Internet Protocol Version 6 (IPv6) orientieren sich am RFC 2460 [10]. Dieses RFC gibt den neuesten Stand der Entwicklung des Internet Protokolls in der neuen Version 6 wieder, denn dieses RFC enthält einige wesentliche Änderungen in der Spezifikation gegenüber dem RFC 1883 [11].

11.3.1 Der IPv6-Header

Die folgende Abbildung zeigt den Header von IPv6 auf, so wie er auch im RFC 2460 [10] dargestellt ist. Im Vergleich zum IPv4 Header ist deutlich die Vereinfachung des Headers zu erkennen. Der Protokollkopf (Header) steuert zu einem grossen Teil die Funktionen des Protokolls, daher wird hier in Abbildung 11.4 zuerst der elementare IPv6-Header (IPv6-Basis-Header) dargestellt. Diese Darstellung wurde mit den Bezeichnungen aus [10] übernommen und zusätzlich mit Grössenangaben in Bits ergänzt. Dieser Header kann jedoch nach Belieben mit verschiedenen Zusatzheadern ergänzt werden, um auf weitere Funktionen zugreifen zu können.

Figure 11.4: IPv6-Basis-Header

Bedeutung der Felder

Version: 4 Bit Internet Protokoll Versionsnummer.

Dieses Feld existiert auch bei anderen Internet Protokoll-Versionen, so zum Beispiel bei IPv4. Somit kann zwischen den verschiedenen Protokollversionen unterschieden werden und es ist ein paralleler Betrieb von IPv4 und IPv6 möglich.

Traffic Class/DS: 8 Bit Feld zur Beschreibung der Priorität.

Dieses Feld wurde aus IPv4 (dortiger Name: Type of Service) mit den dortigen Bedeutungen übernommen. In diesem Feld kann die Absenderapplikation angeben, mit welcher Priorität die Pakete auf dem Weg zum Empfänger behandelt werden sollen. Diese Einstellungen sollen aber nicht vom Benutzer bewusst verändert werden können um sich einen Vorteil daraus zu ziehen, sondern müssen vom System, in einer wohl definierten Gesamtstrategie, geschehen. Durch dieses Feld haben die Router des Transportweges die Möglichkeit, je nach Netzauslastung und Belegung des Class Feldes, die Pakete auf unterschiedlichen Wegen zum Empfänger zu senden, bei Netzwerkstaus die Pakete zu verwerfen oder bei Bedarf die Werte des Class Feld zu verändern. Zurzeit arbeitete eine IETF Arbeitsgruppe daran, dieses Feld genau zu definieren, da es noch nicht vollständig spezifiziert wurde. Fest steht jedoch, dass dieses Feld in IPv4 und IPv6 enthalten sein wird. RFC2474 [12] und RFC2475 [13] befassen sich mit der Definition dieses Feldes, das nun neu Differentiated Service Code Point (DSCP) genannt werden soll.

Flow Label: 20 Bit Identifikationsnummer.

Dieses Feld wird gebraucht um Pakete zu kennzeichnen, für die eine besondere Behandlung erwünscht wird. Es kann also dazu verwendet werden, um zusammengehörende Datenflüsse wie z. B. Video- oder Audiodaten speziell zu behandeln. Falls keine besondere Behandlung des Pakets erwünscht wird, beträgt der Wert des Feldes null. Dadurch entsteht die Möglichkeit, zusammengehörende Pakete zu kennzeichnen und gesondert zu behandeln. In IPv4 musste diese Analyse auf einer höheren ISO/OSI-Schicht vorgenommen werden, was vor allem bei verschlüsselten TCP- oder UDP-Headern nicht lösbar war. Daher ist das Flow-Feld immer unverschlüsselt und steht ganz vorne im Header.

Payload Length: 16 Bit unsigned Integer.

Dieser Wert gibt die Grösse des Paketes nach dem IPv6-Header an. Diese in Bytes angegebene Grösse entspricht der Anzahl der Nutzdaten im Paket, wobei eventuelle nachfolgende, optionale Header in diesem Wert eingeschlossen sind.

Next Header (NH): 8 Bit Selector.

Dieser Wert identifiziert den Typ des ersten Erweiterungsheaders oder eines Layer 4 Headers nach dem Basisheader. Hier wurde zum Teil das Konzept von IPv4 übernommen, jedoch durch weitere optionale Erweiterungsheader wesentlich erweitert. So wird beim Transport eines TCP-oder eines UDP-Pakets sowohl in IPv4 als auch in IPv6 die gleiche Kodierung verwendet.

Hop Limit: 8 Bit unsigned Integer.

Dieser Wert gibt die maximale Anzahl von Vermittlungsrechnern bzw. Routern auf dem Weg des Paketes an. In jedem Knoten, der vom Paket durchlaufen wird, wird dieses Feld um 1 dekrementiert und bei Erreichen des Wertes 0 wird das gesamte Paket verworfen, da es sein Hop Limit erreicht hat. Der Absender erfährt den Verlust des Paketes durch eine automatisch generierte ICMP-Nachricht, die ihm zugesandt wird. Dieses Feld wird dazu benötigt, die Unerreichbarkeit von Knoten zu erkennen und die Pakete nicht über eine längere Dauer im Netz zirkulieren zu lassen. Dieses Feld entspricht ganz dem Time-To-Live-Feld (TTL) im IPv4 Protokoll. Als Startwert kann der Absender Werte von 1 bis 255 eintragen, wobei die Standarteinstellung bei 64 Knoten liegt.

Source Address (SA): 128 Bit lange IPV6-Quelladresse.

Dieses Feld dient zur Identifizierung der Quelladresse, also des Absenders des Paketes. Im Gegensatz zu IPv4 werden hier statt 32 Bit 128 Bit Adressen verwendet. Eine genauere Beschreibung der IPv6-Adressen erfolgt im kommenden Abschnitt.

Destination Address (DA): 128 Bit lange IPV6-Zieladresse.

Dieses Feld dient zur Identifizierung der Zieladresse, also des Empfängers des Paketes. Im Gegensatz zu IPv4 werden hier statt 32 Bit 128 Bit Adressen verwendet. Eine genauere Beschreibung der IPv6-Adressen erfolgt im kommenden Abschnitt.

Erweiterungsheader

Zurzeit sind die folgenden Erweiterungsheader definiert. Das Protokoll bietet aber die Möglichkeit,auch noch weitere Header aufzunehmen.

Hop-By-Hop Header In diesem Header werden Optionen transportiert, welche bei jedem Transportschritt ausgewertet werden müssen.

Routing Header Durch Eintragungen in diesem Header kann der Weg des Paketes zum Zielrechner entscheidend beeinflusst werden. Der Absender kann in diesem Header angeben, über welche Knoten das Paket laufen soll. In der aktuellen Protokollversion wird nur das "'loose routing"' unterstützt, dabei darf das Paket, im Gegensatz zum "'strict routing"', auch über Knoten laufen, welche nicht explizit angegeben sind.

Fragment Header In IPv4 war es möglich, Pakete unterwegs zum Zielrechner zu fragmentieren. Dies ist in IPv6 nicht mehr möglich, der Absender muss also die maximale Paketgrösse auf dem Transportpfad bestimmen und die Pakete selbst fragmentieren. Dieser Header enthält Informationen mit denen der Zielrechner die Pakete wieder in der richtigen Reihenfolge zusammensetzen kann.

Authentification Header Dieser Header erlaubt dem Empfänger festzustellen, ob das Paket tatsächlich von dem angegebenen Absender stammt oder ob es während der Übertragung verändert wurde.

Encapsulation Security Payload Header Um Pakete verschlüsseln zu können, wird dieser Header benötigt.

Destination Options Header In diesem Header können optional Informationen transportiert werden, welche ausschliesslich vom Zielrechner ausgewertet werden.

Wird mehr als nur ein Erweiterungsheader verwendet, ist die Reihenfolge der Erweiterungsheader wie gemäss der obigen Vorstellung einzuhalten. Ausser dem Destination Header darf jeder Header nur einmal verwendet werden.

11.3.2 IPv6-Adressen

Wie bereits angetönt findet sich die grösste Veränderung zwischen IPv4 und IPv6 in der Länge der Adressen. Dadurch vergrössert sich der Adressraum, aber auch die Struktur der Adressen ermöglicht ein optimales Routing und einen flexiblen Aufbau von Adresshierarchien.

Adressformat

Die Schreibweise von IPv6 Adressen unterscheidet sich sehr stark zu den IPv4 Adressen. Der Adressbereich von 128 Bit wurde in jeweils acht 16 Bit lange Adressabschnitte unterteilt. Diese Abschnitte sind durch einen Doppelpunkt voneinander getrennt. Durch die grössere Länge der IPv6 Adressen hat man sich entschieden, die 16 Bit Integerzahl durch ein Tupel von Heydadezimal-Zahlen darzustellen. Eine gültige IPv6 Adresse wäre z.B. folgende: *4030:00BC:0000:00A4:0267:01FF:FE01:7352* Um die Länge der Adresse zu verringern, dürfen die führenden Nullen innerhalb eines Blockes auch weggelassen werden. Ein Block muss aber mindestens eine Hexadezimalziffer enthalten. Die obige IPv6 Adresse kann man folglich wie unten dargestellt schreiben: *4030:BC:0:A4:267:1FF:FE01:7352* Die Länge konnte durch diese Schreibweise jedoch nicht stark verkleinert werden. Da IPv6 Adressen oft lange Null-Folgen enthalten wurde eine weitere Regel eingeführt. Es kann innerhalb einer IPv6 Adresse maximal einmal eine Nullfolge abgekürzt werden, indem die Nullen gar nicht geschrieben werden. Der zweimal hintereinander vorkommende Doppelpunkt steht für die an dieser Stelle fehlende Nullfolge.

Beispiel:

FE80:0:0:0:0:0:0:57 kann auf FE80::57 abgekürzt werden, oder

FE80:0:0:2:0:0:0:5 kann auf FE08:0:0:2::5 abgekürzt werden.

Die Erweiterungsregel lautet:

Der linke Teil der Adresse steht links von den beiden Doppelpunkten, der rechte Teil der Adresse rechts von beiden Doppelpunkten. Dazwischen wird mit Nullgruppen auf 8 jeweils 4-stellige Hexadezimalzahlen aufgefüllt. Bei den

4-stelligen Hexadezimalzahlen dürfen Nullen weggelassen werden, es muss aber mindestens eine Ziffer vorhanden sein. Die Abkürzung für eine Nullgruppe durch zwei aufeinander folgende Doppelpunkte (::) darf aus Eindeutigkeitsgründen nur einmal vorkommen. [14]

Adresstypen

In IPv6 sind drei Typen von Adressen definiert. Dies sind die Typen Unicast, Multicast und Anycast, die im nachfolgenden Abschnitt kurz erläutert werden. Durch den differenzierteren Einsatz von Multicast-Adressen in IPv6 konnte der Einsatz von Broadcast-Adressen wie sie im IPv4 vorhanden sind, ersetzt werden.

- **Unicast**

  Eine Unicast-Adresse dient zur Adressierung eines Interfaces (Schnittstelle) eines IP-Netzwerkknotens, wobei ein Interface in der Regel mehrere Internetadressen hat. Ein Knoten kann aber über jede seiner Unicast-Adressen adressiert werden. Diese Punkt-zu-Punkt Adressierungsart ist somit die häufigste verwendete Art und gilt weltweit. Unicast-Adressen können an den ersten 3 Bits, die den Wert 010 haben, erkannt werden. Eine besondere Bedeutung haben die beiden folgenden Unicast-Adressen:

  0:0:0:0:0:0:0:1

  Mit dieser Adresse ist der eigene Rechnerknoten gemeint und wird Loopback-Adresse genannt. Dies ist genau wie die Adresse 127.0.0.1 in IPv4 eine Pseudoadresse, wo keine Pakete an das Netzwerk verschickt werden, der Knoten verhält sich jedoch so als ob die Pakete an ihn verschickt würden. Diese Adresse kann durch die Möglichkeit der verkürzten Schreibweise auch als ::1 geschrieben werden.

  0:0:0:0:0:0:0:0

  Kennt der Absender seine IP-Adresse noch nicht, fügt er die obige unspezifizierte IP-Adresse als Absender-Adresse ein. Die Abgekürzte Schreibweise besteht lediglich aus zwei Doppelpunkten "'::"'.

- **Multicast**

  Die Multicast-Adressen dienen zur Adressierung einer Gruppe von Interfaces, typischerweise in verschiedenen Netzwerkknoten und können nicht als Absender-Adressen eingesetzt werden. Sie sind wie die Unicast-Adressen durch besondere Adressbereiche erkennbar. Ein IP-Paket, das an eine Multicast-Adresse geschickt wird, wird an alle Mitglieder dieser Multicast-Gruppe gesendet.

- **Anycast**

  Dieser Adresstyp wurde im IPv6 Protokoll neu eingeführt. Versendet man eine Nachricht an eine Anycast-Adresse, erreicht man genau ein Interface aus der ganzen Gruppe. In der Regel wird dies der Knoten sein, der dem Absenderknoten am

Table 11.2: Aufteilung des IPv6-Adressraumes [14]

Adressbereichs-Einteilung	**Adresspräfix**	**Anteil in Prozent**
reserviert für spezielle Anwendungen	00	0.4
nicht zugewiesen	01	0.4
Reserviert für NSAP-Adressen	02 .. 03	0.8
Reserviert für IPX-Adressen	04 .. 05	0.8
nicht zugewiesen	06 .. 07	0.8
nicht zugewiesen	08 .. 0F	3.1
nicht zugewiesen	10 .. 1F	6.3
Aggregierbare globale Unicast-Adressen	20 .. 3F	12.5
nicht zugewiesen	40 .. 5F	12.5
nicht zugewiesen	60 .. 7F	12.5
Geografisch eingeteilte Unicast-Adressen	80 .. 9F	12.5
nicht zugewiesen	A0 .. BF	12.5
nicht zugewiesen	C0 .. DF	12.5
nicht zugewiesen	E0 .. EF	6.3
nicht zugewiesen	F0 .. F7	3.1
nicht zugewiesen	F8 .. FB	1.6
nicht zugewiesen	FC	0.8
nicht zugewiesen	FE 00 .. FE 7F	0.2
Verbindungslokale Netzadressen (link-local)	FE 80 .. FE BF	0.1
Netzlokale Netzadressen (site-local)	FE C0 .. FE FF	0.1
Multicast Adressen	FF	0.4

nächsten liegt, respektive der Knoten, der als erster antwortet. Dieser Adresstyp ist vor allem für mobile Anwendungen äusserst praktisch. Beispielsweise können Name-Server in einer Anycast-Gruppe zusammengefasst werden. Die DNS-Anfragen (Domain Name Service) beantwortet dann der Name-Server, der am schnellsten erreichbar ist. Anycast-Adressen unterscheiden sich nicht von Unicast-Adressen, müssen also beim Konfigurieren explizit als solche deklariert werden, ansonsten wird die Adresse vom Router automatisch als doppelt vergebene Adresse erkannt. Wie bei den Multicast-Adressen, kann eine Anycast-Adresse nie Quell-Adresse eines Paketes sein.

Struktur von Adressen

In ihrer allgemeinen Form sind die IPv6 Adressen nicht strukturiert. Eine Strukturierung kann aber über so genannte Format-Präfixe erfolgen. Das Präfix umfasst die führenden Bits der IPv6 Adresse und besitzt eine variable Länge. Dadurch wird auch der Adresstyp bestimmt. Im Gegensatz zu IPv4 ist nun eine deutlich flexiblere und aufwendigere Unterscheidung der Internet-Adressen realisiert. Aus Tabelle 11.2, die die Verteilung der Adressbereiche aufzeigt, ist ersichtlich, dass lediglich 27.6 Prozent der möglichen Adressbereiche in Benutzung sind.

11.3.3 Routing in IPv6

Hauptaufgabe des Internetprotokolls ist die Übertragung von einem Endsystem (Sender) über ein komplexes Netz zu einem Partner-Endsystem (Empfänger). Zwischen Absender und Empfänger befinden sich mehrere Knoten, über die die Nachricht gesendet wird. Es muss also jeder Knoten dazwischen wissen, wohin er das Paket weiterleiten muss, damit es schlussendlich beim Empfänger ankommt. Die Knoten, die eine solche Wegauswahl treffen werden Router genannt und können mit Hilfe von Routing-Algorithmen herausfinden, welcher der beste Weg für ein Paket an die Zieladresse ist. Bei der Wegauswahl ist der Aufbau der IP-Adressen von entscheidender Bedeutung, daher wurde beim Entwurf der Adressarchitektur Wert darauf gelegt, dass die Wegauswahl relativ leicht vorgenommen werden kann. Bei allen Routing-Entscheidungen wird das Paket an das Netz weiter geleitet, dessen Adress-Präfix die längste Übereinstimmung mit dem Adress-Präfix des IP-Paketes vorliegt. Beispielsweise haben wir ein Netz 1 mit der Präfix 2ABC:: und ein Netz mit dem Präfix 2ABC:0034::. Ein IP-Paket mit dem Ziel 2ABC:0034:5678:3::25 wird an Netz 2 weitergeleitet, da das IP-Paket mit dem Netz 2 die längere Übereinstimmung im Adres-Präfix hat als mit Netz 1. In IPv6 gibt es aggregierbare Adressen. Diese ersetzen die von IPv4 her bekannten öffentlichen Adressen (IPv4 Class A, B und C). Für eine Adressaufteilung nach Providern ist dieser Adressbereich nochmals feiner unterteilt worden. Diese feinere Aufteilung ist in Abbildung 11.5

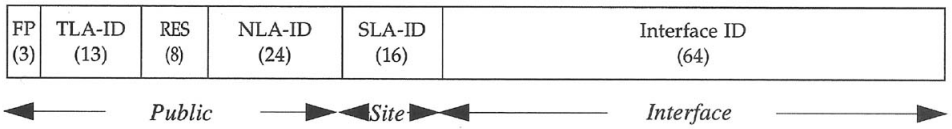

Figure 11.5: Aufteilung einer aggregierbaren Adresse

FP: Format Präfix Der Format-Präfix zeigt den Bereich der aggregierbaren globalen Unicast-Adressen an, dessen Wert in Tabelle 11.2 nachgesehen werden kann.

TLA-ID: Top-Level Aggregation Identifier

Hier wird ein nationaler oder internationaler grosser Provider der obersten Hierarchiestufe eines Netzproviders gekennzeichnet.

RES: Reserviert

Dieses Feld ist für zukünftige Anwendungen reserviert und muss im Moment den Wert 0 besitzen. Das Reservefeld ist für Vergrösserungen der Felder TLA-ID oder NLA-ID reserviert, was jedoch in der heutigen Zeit absolut nicht nötig ist.

NLA-ID: Next-Level Aggregation Identifier

Hier wird ein regionales Netz oder ein Teilnetz eines grossen Providers gekennzeichnet. Dieses Feld kennzeichnet also die nächstniedrige Hierarchiestufe nach dem Top-Level Kennzeichner (TLA-ID). Ein Provider besitzt auch die Möglichkeit dieses Feld selbst weiter aufzuteilen.

SLA-ID: Site-Level Aggregation Identifier

Dieses Feld gehört dem angeschlossenen privaten Netzbetreiber, das zum Beispiel ein Firmennetzwerk oder das Netz einer Hochschule sein kann. Der private Netzbetreiber kann dieses Feld zur Bildung von Subnetzen verwenden.

Interface ID: Interface Identifier

Dient zur Kennzeichnung der Schnittstelle eines Netzknotens. Dieses Feld muss bezüglich der unterliegenden OSI-Ebene-2 innerhalb des Link Segments eindeutig sein, das heisst in diesem Adressraum darf diese Adresse nur genau einmal vorkommen. Der Interface ID wird häufig auch als MAC-Adresse bezeichnet, welches die Link-Layer-Adresse (OSI ebene 2) der Schnittstelle ist. Diese 64 Bit Interface ID können durch eine feste Regel aus den 48 Bit MAC-Adressen generiert werden.

Zusammenfassend kann gesagt werden, dass der öffentliche Teil einer IPv6-Adresse aus den Feldern FP bis NLA-ID gebildet wird. Diese Felder werden weltweit eindeutig vergeben. Der TLA-ID identifiziert eindeutig einen privaten Netzbetreiber. Dieser kann seinen NLA-ID in einer oder mehreren Hierarchiestufen nach regionalen oder organisatorischen Gesichtspunkten aufteilen.

Beim Routing muss weiter zwischen dem Routen von Paketen innerhalb eines Privaten Netzwerks (Intra Domain Routing) und zwischen dem Routen von Pakten im Internet (Inter Domain Routing) unterschieden werden. Durch die unterschiedlichen Anforderungen an die beiden Routing-Varianten, werden unterschiedliche Protokolle mit unterschiedlichen Eigenschaften eingesetzt. Daher sollen nun kurz die Anforderungen an die Routingalgorihtmen im Inter- und Intra Domain Routing aufgezeigt werden.

Intra Domain Routing

Das Intra Domain Routing bezeichnet das Routing innerhalb eines privaten Netzwerks. Dieses besitzt im Vergleich zum weltweiten Internet viel weniger Knoten, die Netzwerktopologie ist viel einfacher aufgebaut und weniger komplex. Aufgrund der geringen Komplexität können beim Intra Domain Routing auch Routingverfahren eingesetzt werden die den Gesamtzustand des Netzes kennen, was bei einem Inter Domain Routing-Verfahren aufgrund der Komplexität und Grösse des Netzes absolut undenkbar wäre. Der OSPF (Open Shortest Path First) Algorithmus ist ein Beispiel für ein Routing Verfahren, das im Intra Domain Routing zur Anwendung kommt. Dieses Verfahren kann sowohl für IPv4 als auch für IPv6 eingesetzt werden und wird in RFC 2328 [15] beschrieben. Für den Betrieb mit IPv6 müssten jedoch die IPv6-Adressen und die IPv6-Präfixe zur Beschreibung des Netzes eingeführt werden, ebenso wird die erweiterte Multicast-Fähigkeit von IPv6 untersützt. Diese und weitere Anpassungen werden in RFC 2740 [16] beschrieben. Bei diesem Algorithmus muss zuerst die gesamte Netzwerk-Topologie bekannt sein. Dies geschieht über den Versand von Hello-Paketen an die Nachbarknoten. Ist die Netzwerk-Topologie bekannt, so kann mit dem Dijkstra-Algorithmus der kürzeste Weg zu einem Zielknoten berechnet werden. [17]

Ein weiteres anwendbares Verfahren ist das RIP (Routing Information Protocol), welches sehr einfach und universal einsetzbar ist. Es ist ein "'Distance Vector"' [17] Verfahren. Das heisst, das Paket wird über den Weg mit den wenigsten Zwischenknoten versendet.

Inter Domain Routing

Für das Inter Domain Routing werden Protokoll aus der Klasse der Exterior Gateway Protocols (EGP) eingesetzt. Bei IPv4 wird vor allem das Border Gateway Protocol (BGP) eingesetzt. Für IPv6 Netze wurde zuerst das Inter Domain Routing Protocol Version2 (IDRPv2) eingesetzt. Mittlerweile wurde das BGP an die aktuelle IPv6 Version angepasst, so dass für IPv6 zwei Routingprotokolle eingesetzt werden könnten. Beide Protokolle sind so genannte Vektor-Pfad-Protokolle, das heisst die Optimierung erfolgt durch die Auswahl des Weges mit den wenigsten Zwischenknoten. Bei dieser Art von Protokollen wird in den Routing-Tabellen nicht nur der nächste Knoten, sondern der ganze Pfad zum Zielknoten gespeichert. Dadurch lassen sich sehr einfach Schleifen erkennen und vermeiden.

11.3.4 Sicherheit in IPv6

In diesem Abschnitt soll auf die im IPv6 Protokoll eingebauten Sicherheitsmechanismen eingegangen werden. Dabei wird auf die Schilderung von verschiedenen Angriffsszenarien verzichtet, da solche zum Teil schon in vorigen Arbeiten behandelt wurden. Wie im Abschnitt "'Der IPv6 Header"' bereits geschildert, sind die Erweiterungsheader, die eine zusätzliche Sicherheit möglich machen, so in der Hierarchie der Erweiterungsheader angeordnet, dass möglichst viele Daten verschlüsselt werden können.

Authentifikation mit dem Authentification Header (AH)

Die Authentifikation von Paketen, also die Sicherheit, dass das Paket auch wirklich von der angegebenen Absenderadresse stammt, wird vom Authentification Header (AH), in Abbildung 11.6 dargestellt, gewährleistet. Dieser Erweiterungsheader kann sowohl in IPv6 als auch in IPv4 verwendet werden.

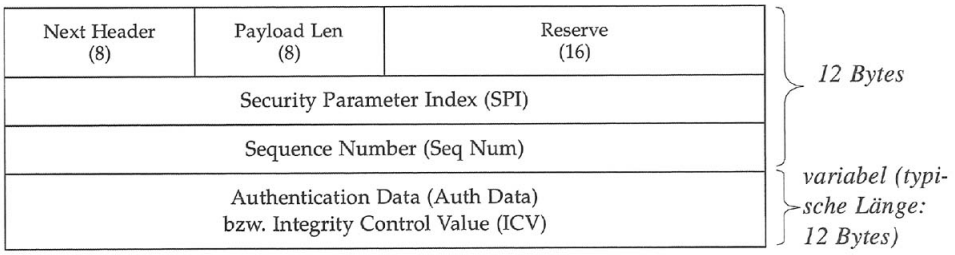

Figure 11.6: Authentification-Header

Next header Legt den Typ des auf den AH folgenden Headers fest.

Payload Len In diesem Feld wird die Länge des gesamten Authentification Headers angegeben, inklusive den Authentifikationsdaten.

Reserve Dieses Feld ist für spätere Erweiterungen reserviert und muss im Moment 0 sein.

SPI (Security Parameter Index) Dieser Wert zeigt auf einen Tabelleneintrag in der Security Association Database des Empfängers und dient damit der Identifikation einer Sicherheitsbeziehung. Die Werte von 1 bis 255 sind reserviert für die weltweit einheitlichen Sicherheitsbeziehungen.

Seq Num Die 32 Bit Zahl in diesem Feld wird monoton um eins erhöht. Hat sie ihren maximalen Wert erreicht, so muss zwischen dem Sender und Empfänger eine neue Sicherheitsbeziehung eingerichtet werden. Die Sequenznummer wird vom Sender erzeugt und dient ausschliesslich dazu, dass Pakete mit identischen Daten nicht wiederholt gesendet werden. Damit können Reply-Attacken verhindert werden.

Auth Data Dieses Feld enthält berechnete Authentifikationsdaten. Die Länge dieses Feldes ist abhängig vom verwendeten Algorithmus, der bei der Einrichtung der Sicherheitsbeziehung vereinbart wurde. Die Daten in diesem Feld werden über folgende Teile eines IPv6-Paketes berechnet:

1. Alle auf dem Transport vom Sender zum Empfänger nicht veränderte Felder des IP-Basis-Headers, sowie geänderte Felder, die vom Empfänger vorhergesagt werden können.
2. Alle Felder des Authentification Headers
3. Eventuelle nachfolgende Erweiterungs-Header sowie über die transportierten Daten der Nutzlast.

Zur Berechnung der Authentifizierungsdaten müssen mindestens die beiden Hash-Funktionen MD5 und SHA-1 unterstützt werden.

Verschlüsselung mit ESP

Mit Hilfe von Authentifizierungsverfahren kann zwar der Absender des Paketes eindeutig identifiziert und authentifiziert werden. Um verhindern zu können, dass der Datenstrom auf dem Weg zum Empfänger von Dritten abgehört wird, muss dieser verschlüsselt werden. Zur Verschlüsselung von Daten existieren zwei Verfahren, die symmetrische und die asymmetrische Verschlüsselung. Diese Verschlüsselung des Datenstroms muss schnell von statten gehen, wobei beachtet werden muss, dass die Geräte zu Weiterleitung nicht eine sehr grosse Rechenleistung besitzen. Daher sollte es auch möglich sein, die Verschlüsselung durch Hardware-Komponenten zu realisieren. Da die asymmetrische Verschlüsselung einen höheren Rechenaufwand verlangt als die symmetrische, ist es angebracht für IP Pakete eine symmetrische Verschlüsselung zu wählen. [18]

In IPv6 ist eine Verschlüsselung von IP Paketen mit Encapsulation Security Payload (ESP) möglich. Dabei wird der ESP-Header vor die zu verschlüsselten Daten gestellt und einen ESP-Trailer dahinter.

Bei der Anwendung der Verschlüsselung kann man je nach Art zwischen Transport- und Tunnelmodus unterscheiden. Um den Unterschied zwischen Transport- und Tunnelmodus zu verdeutlichen, nehmen wir als Beispiel das in Abbildung 11.7 dargestellte Original IP-Paket ohne Verschlüsselung.

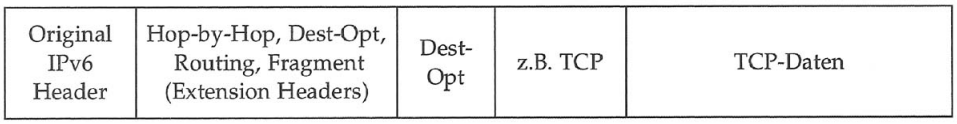

Figure 11.7: Original IP-Paket

Nach den optionalen Erweiterungsheader folgt als zweitletzter Erweiterungsheader der ESP-Header, die Daten im Paket werden verschlüsselt und authentifiziert. In Abbildung 11.8 ist dargestellt, welche Daten verschlüsselt und welche auch noch mit ESP-authentifiziert werden.

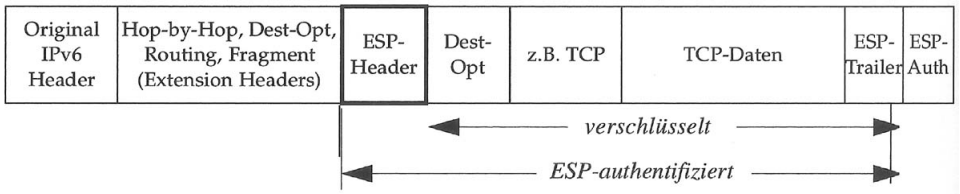

Figure 11.8: IP-Paket im Transportmodus

Im Tunnelmodus wird das verschlüsselte Paket nicht vom Sender verschlüsselt und vom Empfänger entschlüsselt, sondern es wird davon ausgegangen, dass sowohl Sender als auch Empfänger sich jeweils in einem sicheren Netz befinden. Für die Kommunikation im eigenen Netz ist es daher nicht nötig, Pakete zu verschlüsseln. Wollen die beiden Rechner jedoch miteinander Daten über sich dazwischen befindendes, unsicheres Netz austauschen, müssen die Daten verschlüsselt werden. Dies geschieht an den so genannten Security-Gateways, welche sich an der Trennlinie zwischen sicherem und unsicherem Netz befindet. Der Security-Gateway fügt um das Original-Paket ein neues IP-Paket ein. Nun kann dieses Paket das unsichere Netz passieren und wird beim Security-Gateway des Empfängers wieder ausgepackt, so dass das Original-Paket an den Empfänger weiter geschickt werden kann. In Abbildung 11.9 ist dieses IP-Paket im Tunnelmodus dargestellt. [14]

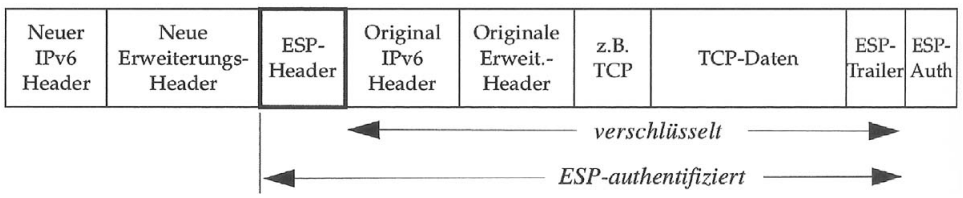

Figure 11.9: IP-Paket im Tunnelmodus

11.3.5 Weitere Funktionen von IPv6

Zustandslose Konfiguration

Bei IPv4 müssen die Adressen der Endgeräte in der Regel manuell eingestellt werden. Dies führte häufig zu Fehleingaben, wie zum Beispiel doppelten IP Adressen, was wiederum den Betrieb des Netzwerkes stören kann. Fallen in einem grossen Netzwerk Änderungen an, so muss jeder Rechner einzeln konfiguriert werden, was ein nicht zu unterschätzender Arbeitsaufwand bedeutet. Nicht zuletzt aus diesem Grund wurde das DHCP-Protokoll (Dynamic Host Configuration Protocol) eingeführt. Mit Hilfe dieses Protokolls ist es möglich, die Netzwerkeinstellungen eines IPv4-Rechners automatisch zu konfigurieren. Dieses Verfahren erfordert aber eine statische Verwaltung von Informationen, was für dynamische Netze ein grosser Nachteil ist. Das DHCP-Protokoll kann auch in IPv6 weiter verwendet werden, doch wurde bei der Entwicklung von IPv6 darauf Wert gelegt, dass das Protokoll eine automatische Konfiguration unterstützt. Die so genannte zustandslose Konfiguration [19] ermöglicht es Geräten, sich in einem Netz selbstständig zu konfigurieren. Dabei entsteht jedoch der Nachteil, dass der Systemadministrator die vergebenen IP Adressen nicht beeinflussen kann. Als erster Schritt bildet die zustandslose Konfiguration eine so genannte link-local Adresse, welche sich aus der Hardwareadresse der Netzwerkkarte und dem Präfix FE80 bis FEBF zusammensetzt. Anschliessend muss mit Hilfe der Duplicate Address Detection verhindert werden, dass eine IP-Adresse doppelt vergeben wird. Sollte dieser Vorgang ergeben, dass eine Adresse doppelt vergeben wurde, so muss das Interface nun manuell konfiguriert werden und die Duplicate Adress Detection erneut durchführen. Wurde keine doppelte IP-Adresse festgestellt, kann der Rechner mit der link-local Adresse im selben Subnetz mit anderen Rechnern kommunizieren. Möchte der Rechner aber in einem globalen Netz kommunizieren, muss dieser zuerst einen Router finden. Dies kann über eine passive oder eine aktive Suche geschehen, oder aber er erlangt diese Informationen über das Finden eines DHCP-Servers. Die globale Adresse wird nun aus dem Präfix und der link-local Adresse erzeugt. Die Gültigkeitsdauer einer IPv6-Adresse ist auf eine bestimmte Zeit begrenzt, das heisst, jedes Gerät muss in regelmässigen Abständen den Prozess der zustandslosen Konfiguration durchlaufen [20].

11.4 Migration von IPv4 auf IPv6

11.4.1 Definition von Migration

Migration ist der periodische Transfer von digitalem Material von einer Hard- /Softwarekonfiguration zu einer anderen Konfiguration, von einer Generation der Computertechnologie zur nachfolgenden [21]. Das Ziel der Migration ist das Erhalten der Integrität von digitalen Objekten. Auf Anwenderseite soll stets gewährleistet sein, daß Daten empfangen, angezeigt oder anders genutzt werden können, trotz sich ständig ändernder Technologien. Diese Definition ist kritisch zu betrachten. Zum Ersten muß das digitale Material nicht periodisch transferiert werden, da Migrationszeitpunkte von sehr vielen Faktoren (z.B. neue Datenstandards, neue Hardware, etc.) abhängen können. Es

kann weiterhin möglich sein, daß Daten nur ein einziges Mal migriert werden. Ein weiterer Kritikpunkt ist das angegebene Ziel, die Integrität von digitalen Objekten zu erhalten. Wird ein digitales Objekt migriert, indem es zum Beispiel in einem neuen Format gespeichert wird, gehen oft Daten des alten Formates verloren oder werden geändert. Formatierungen werden oftmals modifiziert oder andere Veränderungen vorgenommen. Für den Übergang auf IPv6 hat man sich daher viele Gedanken über eine sanfte Migration gemacht. Innerhalb der Internet Engineering Task Force gründete man hierfür eine eigene Arbeitsgruppe. Diese soll verschiedene Modelle für den Übergang und die Einführung ausarbeiten. In RFC1933 wurde ein Mechanismus definiert, der für den sanften Umstieg von einer Protokollversion zur nächsten sorgt [22]. DNS-Server müssen mit den in RFC1886 definierten Erweiterungen auf IPv6 vorbereitet sein [23].

Migrationstechniken

Es ist nicht möglich weltweit das IPv4 Protokoll in einem Zug auf das IPv6 Protokoll umzustellen. Ein wesentliches Ziel bei der Entwicklung von IPv6 war es, eine möglichsteinfache und fließende Migration von IPv4 zu IPv6 zu gewährleisten. Hierzu wurden Transitions Mechanismen für IPv6-Hosts und -Router definiert [24]. In den folgenden Abschnitten werden Techniken vorgestellt, welche für eine Migration von IPv4 zu IPv6 entworfen wurden.

11.4.2 Dual Stack

Eine Kompatibilität zur bestehenden IPv4-Infrastruktur ist besonders in der Anfangsphase von IPv6 nötig. So werden IPv4-Protokolle genutzt werden müssen, wenn zwei IPv6-Knoten nicht direkt über eine IPv6-Infrastruktur miteinander verbunden sind, aber dennoch miteinander kommunizieren wollen. Um dies zu ermöglichen kann eine Dual-IP-Layer-Implementation auf den Knoten eingesetzt werden. Die Dual-IP-Layer-Technik wird als Dual-Stack bezeichnet.

Dual-Stack bedeutet, dass auf den Knoten jeweils eine eigenständige Implementation von IPv4 und von IPv6 zur Verfügung steht, d.h. es ist sowohl IPv4 als auch IPv6 als Internet-Protokoll implementiert.

Diese Knoten, auch als IPv4/IPv6-Knoten bezeichnet, können sowohl IPv4- als auch IPv6-Pakete senden und empfangen. Folglich können IPv4/IPv6-Knoten mit IPv6-Knoten über IPv6 und mit IPv4-Knoten über IPv4 kommunizieren. Das Routen der Pakete erfolgt dabei mittels des jeweiligen Routing-Protokolls.

11.4.3 Tunneling

Die Dual IP-Layer-Technik kann in Verbindung des Tunneling von IPv6-Paketen über IPv4-Netze angewendet werden. Damit können IPv6-Geräte über IPv4 miteinander Daten austauschen. Die IPv6-Datenpakete werden dazu in IPv4-Datenpakete eingebunden.

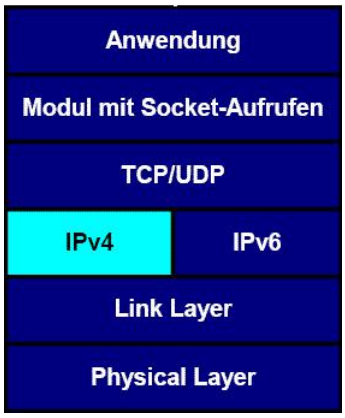

Figure 11.10: Schematische Darstellung der Dual Stack Architektur am TCP/IP-Schichten Modell

Table 11.3: Durch Dual IP-Layer mögliche Kommunikationsszenarien und das verwendete Protokoll

Source	**Destiantion**	**Protokoll**
IPv4/IPv6-Knoten	IPv6-Knoten	IPv6
IPv4/IPv6-Knoten	IPv4-Knoten	IPv4
IPv6-Knoten	IPv4/IPv6-Knoten	IPv6
IPv4-Knoten	IPv4/IPv6-Knoten	IPv4

Dazu wird das gesamte IPv6-Paket als Payload-Information auf das IPv4-Paket gepackt. Natürlich kann das Verfahren auch umgekehrt angewendet werden.

Ein IPv4/IPv6 Knoten der das Tunneling unterstützt, kann entweder ausschließlich manuell konfiguriertes Tunneling unterstützen oder beides - manuell konfiguriertes und automatisches Tunneling. Es gibt demnach drei Implementierungsmöglichkeiten für einen IPv6/IPv4-Knoten: [25]

a) es wird kein Tunneling unterstützt.

b) Tunneling erfolgt ausschließlich über manuell voreingestellte Tunnelpfade.

c) sowohl manuell konfiguriertes, als auch automatisches Tunneling ist möglich.

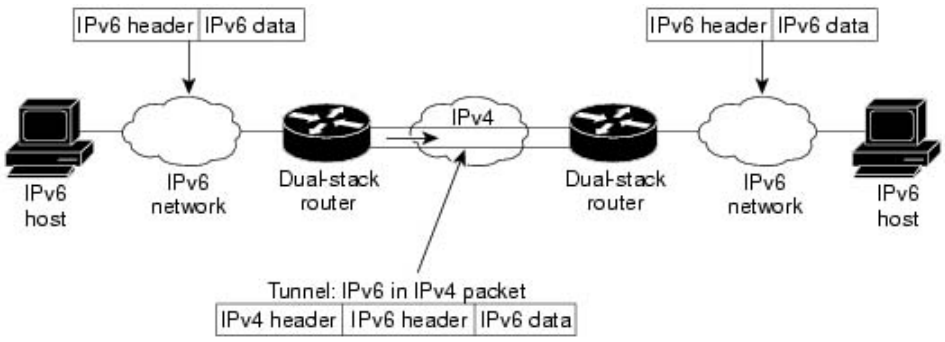

Figure 11.11: Schematische Darstellung von Tunneling über Stacks

Sonderadressen für Übergang von IPv4 zu IPv6

Um eine fließende Migration von IPv4 zu IPv6 zu gewährleisten, wurden unter anderem folgende Adressen definiert:

Table 11.4: Sonderadressen für Migration von IPv4 zu IPv6

Sonderadresse	**Adress-Präfix**	**Adressbildung**	**Bedeutung**
6to4-Adresse	2002/16	2002 + IPv4Addr+::+ Interface-ID	automatisches Tunneln von IPv4-Paketen über ein IPv6-Netz
IPv4-kompatible IPv6-Adresse	::/96	:: + IPv4Addr	Übertragung von IPv6- Adressen über ein IPv4-Netz
IPv4-mapped IPv6-Adresse	::FFFF/96	:FFFF + IPv4Addr	Übertragung von IPv4- Adressen über ein IPv6-Netz
IPv4-translated IPv6-Adresse	::FFFF:0/96	::FFFF:0 + IPv4Addr	Adresse wird automatisch in IPv4- bzw. IPv4-translated IPv6-Adresse umgewandelt

6to4-Adressen sind in RFC3056 [26] beschrieben und werden verwendet, um IPv6-Pakete automatisch über ein IPv4-Netzwerk zu tunneln. Die Adresse setzt sich aus dem Präfix 2002::/16, der sich aus dem Format Präfix 001 und dem TLA-Wert 2 welcher für 6to4-Systeme reserviert ist bildet, sowie einer anschließenden 32 Bit IPv4-Adresse und den abschließenden 64 Bit Interface-Identifier zusammen (2002:IPv4-Adresse::/48).

Bei einer IPv4-kompatible IPv6-Adresse handelt es sich um eine Adresse eines ansonsten IPv6 fähigen Hosts. Sie wird genutzt, um IPv6-Adressen in einem IPv4 zu einem anderen IPv6 Netzwerk zu transportieren, wobei ebenfalls Tunnel verwendet werden. Die oberen 96 Bit der Adresse sind Null,woraus sich der Adresspräfix ::/96 ergibt. Die restlichen 32 Bit werden durch die IPv4-Adresse gefüllt (::IPv4-Addresse). IPv4-abgebildete IPv6-Adressen werden für die Kommunikation zwischen einem IPv6/IPv4-Host und einem IPv4-Host genutzt. Die Adresse setzt sich aus dem Adresspräfix ::FFFF/96 und einer IPv4-Adresse zusammen. Eine IPv4 konvertierte IPv6-Adresse wird von einem Protokollübersetzer erkannt und automatisch in eine IPv4-Adresse oder auch umgekehrt umgewandelt. Die unteren 32 Bit der Adresse bestehen aus einer IPv4-Adresse und die oberen 96 Bit bilden sich aus dem Adresspräfix ::FFFF:0/96.

IPv6-over-IPv4 Tunneling

In der Einführungsphase von IPv6 in eine bestehende IPv4-Umgebung werden IPv6-Netzwerke innerhalb des IPv4-Netzes implementiert. Hierdurch entstehen sogenannte IPv6-Inseln, die miteinander über das IPv4-Netz kommunizieren.

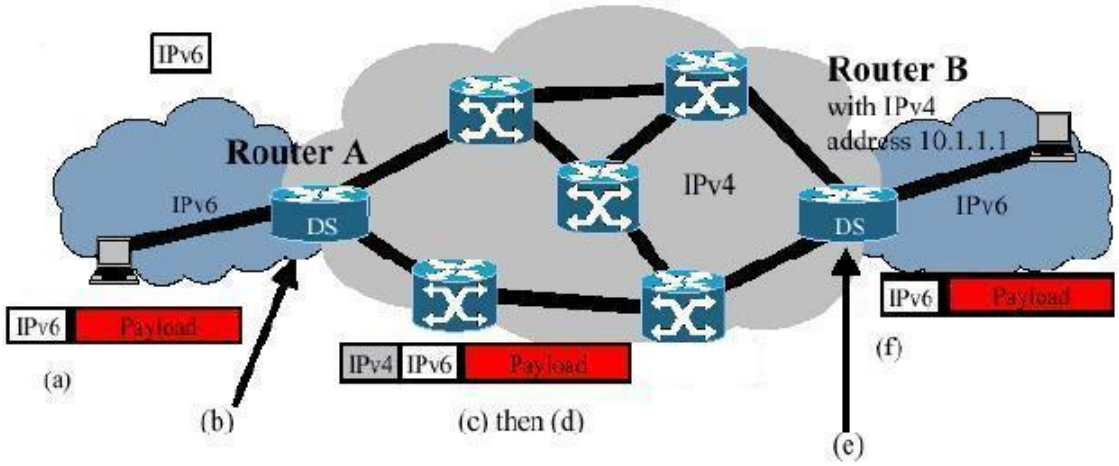

Figure 11.12: Darstellung des Ablaufs beim IPv6-over-IPv4-Tunneling

(a) Ein Paket mit einer IPv6-Zieladresse erreicht Router A

(b) Router A erkennt anhand seiner Routingtabelle, dass das Paket an Router B weitergeleitet werden muss. Ebenfalls findet Router A die IPv4-Adresse 10.1.1.1 von Router B heraus (konfigurierte oder automatische Tunnel).

(c) Das IPv6-Paket ist in ein IPv4-Paket gekapselt und wird über das IPv4- Netz an Router B gesendet.

(d) Das IPv4-Netz routet das Paket mit der Zieladresse 10.1.1.1, als wäre es ein IPv4-Paket, bis zum Router B.

(e) Router B schaut sich das Paket an und erkennt, dass eine IPv6-Paket enthalten ist. Er entfernt den IPv4-Header und nutzt den IPv6-Header um die IPv6-Zieladresse mit seiner Routingtabelle zu vergleichen. Er erkennt, dass das Ziel innerhalb seines Netzes liegt.

(f) Router B sendet das IPv6-Paket zu seinem Ziel

Wie eben dargestellt, tunneln IPv4/IPv6-Hosts und -Router IPv6-Pakete über ein IPv4-Netz, indem die IPv6-Pakete in ein IPv4-Datagram gekapselt werden. Neben dem dargestellten Szenario wäre es auch denkbar, dass nur ein einzelner IPv4/IPv6-Host in einer IPv4-Umgebung vorkommt. Hierbei gibt es keinen Router der die Pakete kapselt bzw. entkapselt, d.h. der Host selber muss die Aufgaben des Tunnel-Endpunktes übernehmen. Bei einem Einsatz von IPv6-over-IPv4 Tunneling wird zwischen konfigurierten und automatischen Tunneln unterschieden.

Konfigurierte Tunnel

Um zwei IPv6-Netze über ein IPv4-Netz durch einen Tunnel zu verbinden, jedoch aus der IPv6-Zieladresse eines Paketes nicht automatisch eine IPv4-Adresse abgeleitet werden kann, so muss ein Tunnel fest konfiguriert werden. Hierfür muss Grundlagen zu IPv6 auf beiden Tunnelendpunkten eine Konfiguration erfolgen. Ebenso müssen beide Endpunkte eine gültige eindeutige IPv4-Adresse besitzen.

Automatische Tunnel

Neben der Möglichkeit Tunnel manuell zu konfigurieren, können Tunnel auch automatisch eingerichtet werden, wenn aus der IPv6-Zieladresse der Tunnelendpunkt abgeleitet werden kann. Damit ein Router erkennen kann an welche IPv4-Adresse das Paket gesendet werden soll, werden IPv4-kompatible- IPv6-Adressen und 6to4-Adressen eingesetzt.

11.4.4 IPv4/IPv6 Protokollübersetzung

Neben dem Einsatz von Dual-IP-Architekturen ist bei der Einführung von IPv6 auch der Fall zu betrachten, dass IPv6-Knoten in einem neu entstandenen reinem IPv6-Netz weiterhin mit IPv4-Knoten, welche sich in einem reinen IPv4-Netz befinden, kommunizieren wollen. Dabei bedeutet reines IPv4- bzw. IPv6-Netz, dass innerhalb dieser Netze nur mit den jeweiligen Netzprotokollen gearbeitet und nur für das jeweilige Netzprotokoll eine Routing Infrastruktur aufgebaut wird.

Um die beschriebenen Kommunikationsszenarien zu ermöglichen, kann eine Protokollübersetzung auf Netzebene (Ebene 3 des OSI-Modells) vorgenommen werden. Hierfür wurden die Mechanismen Stateless IP/ICMP Translation (SIIT) [27] und Network Address Translation - Protocol Translation (NAT-PT) [28] definiert.

Bei beiden Mechanismen werden IPv4-Header in IPv6-Header und umgekehrt transformiert. Dabei können aufgrund der erheblichen Unterschiede zwischen den beiden Protokollversionen die folgenden Felder nicht übersetzt werden:

1. Alle IPv4-Optionen im IPv4-Header.
2. IPv6-Routing-Header.
3. IPv6-Hop-by-Hop-Erweiterungsheader.
4. IPv6-Destination-Options-Header.

Von den Erweiterungsheadern für IPSec ist nur der ESP-Header im Transport-Mode transformierbar. ESP im Tunnel-Mode und der Authentification-Header AH sind nicht übersetzbar, da einige Felder, welche übersetzt werden, in die Berechnung der verschlüsselten bzw. authentifizierten Daten einbezogen sind. Des Weiteren sind von den ICMP-Nachrichten nur die ICMP-Typen transformierbar, welche in beiden Protokollversionen vorhandenen sind. Die für beide IP-Versionen elementaren Felder wie z.B. Quell- und Zieladresse oder Next-Header bzw. Protokoll sind in beide Richtungen übersetzbar.

Stateless IP/ICMP Translation (SIIT)

Mit SIIT wird eine zustandslose Protokollübersetzung von IP- und ICMP-Protokolle bezeichnet. Dabei bedeutet zustandslos, dass jedes Paket für sich, ohne Speicherung eines Kontextes, übersetzbar ist.

Die Voraussetzung für SIIT ist eine dynamische Allokation von IPv4-Adressen. Um dies zu ermöglichen werden IPv4-translated IPv6-, IPv4-mapped IPv6- und IPv4- kompatible IPv6-Adressen verwendet. Neben der Allokation der IPv4-Adresse ist als weitere Voraussetzung für eine Kommunikation eine öffentliche IPv4-Adresse für den IPv6-Host bereitzustellen.

Die Aufgaben bzw. Funktionen der Protokollübersetzung werden in einer so genannten SIIT-Box bzw. im Translator, einem mit Zusatzsoftware ausgestattetem Router, übernommen. Die SIIT-Box bzw. der Translator ist am Rand eines IPv6- Netzes zum Übergang an ein IPv4-Netz stationiert.

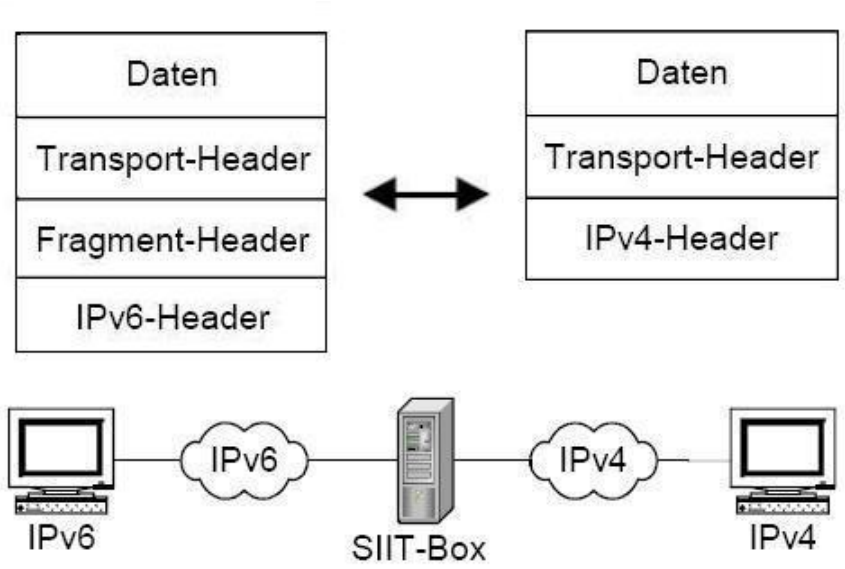

Figure 11.13: schematische Darstellung von SIIT zwischen einem IPv6-und einem IPv4-Host über eine SIIT-Box

11.4.5 Network Address Translation - Protocol Translation (NAT-PT)

Mit NAT-PT wird die Voraussetzung einer IPv4-Adresse für den IPv6-Host in SIIT übergangen. Hierbei wird der aus IPv4 bekannte Mechanismus Network Address Translation (NAT) mit der Protokollübersetzung gemäß SIIT kombiniert.

Der Vorteil gegenüber SIIT ist, dass für eine Kommunikation zwischen einem IPv6- und einem IPv4-Host für den IPv6-Host keine IPv4-Adresse mehr bereitgestellt werden muss, was in Anbetracht einer Adressknappheit von großem Nutzen sein kann.

Nachteil gegenüber SIIT ist, dass es sich bei NAT-PT um keine zustandslose Protokollübersetzung mehr handelt. Denn die Identifikation einer Kommunikationsbeziehung wird bei Einrichten einer Verbindung im NAT-PT Gateway gespeichert und für eine spätere Protokollübersetzung gebraucht. Hierfür müssen größere Rechen- und Speicher-Ressourcen auf dem NAT-PT-Gateway als für eine SIIT-Box bereitgestellt werden.

Das bekannte Problem von NAT, dass eine Verbindung nur von einem Host innerhalb des mit NAT betrieben Netzes aufgebaut werden kann, wird durch die als Bi-Directional-NAT-PT bezeichnete Betriebsart von NAT-PT umgangen. Bei der Realisierung eines bidirektionalen NAT-PT wird das DNS in Anspruch genommen. Der IPv6-Adresse des DNS-Servers im IPv6-Netz wird eine IPv4-Adresse permanent zugeordnet.

Bei Bi-Directional NAT-PT sind zwei Situationen zu unterscheiden:

a) Nur ein IPv6-Netz wird mit einem IPv4-Netz vernetzt.

b) Zwei IPv6-Netzwerke werden über das Internet vernetzt.

Damit das Prinzip von Bi-Directional NAT-PT vollständig dargestellt werden kann, werden folgende Fälle beim Initiieren der Kommunikation betrachtet:

a) IPv6-Netz (IPv4-Netz: Die Kommunikation initiiert ein Rechner im IPv6-Netz zu einem Rechner im IPv4-Netzwerk.

b) IPv4-Netz (IPv6-Netz: Die Kommunikation initiiert ein Rechner im IPv4-Netz zu einem Rechner im IPv6-Netzwerk.

c) IPv6-Netz (IPv6-Netz über das Internet: Die Kommunikation initiiert ein Rechner in einem IPv6-Netz zu einen Rechner in einem anderen IPv6-Netz, wobei die beiden IPv6-Netzwerke über das Internet vernetzt sind.

Hierdurch ist es möglich, auch eine Kommunikation aus einem IPv4-Netz in ein IPv6-Netz aufzubauen.

11.4.6 Transport Relay Translator (TRT)

Neben einer Protokollübersetzung auf Netzebene, wie sie im vorangegangen Abschnitt beschrieben wurde, bietet TRT die Möglichkeit einer Kommunikation zwischen reinen IPv6- und reinen IPv4-Netzen, durch eine Protokollübersetzung auf Transportebene (Ebene 4 des OSI Modells.

Diese Technologie ist vor allem für die Fälle vorgesehen, in denen IPv6-Knoten Dienste von IPv4-Servern in Anspruch nehmen wollen. Es wurde dafür ein so genannter IPv6-to-IPv4-Transport-Relay-Translator (TRT) definiert [29].

Bei TRT findet eine Übersetzung von TCP/UDP-IPv6 nach TCP/UDP-IPv4 und umgekehrt statt. Hierbei können sich die Kommunikationspartner in reinen IPv6- bzw. IPv4-Netzen befinden, dass heißt es ist nur eine Routing-Infrastruktur für die entsprechende IP-Version notwendig.

Voraussetzung für TRT ist ein Modifizierter Name-Server, der netzlokale Pseudo- IPv6-Adressen mit Präfix des TRT-Systems erzeugt, welches diese Adresse bei Erhalt von Paketen durch IPv4-Adressen ersetzt.

Der wesentliche Vorteil dieses Verfahrens ist, dass keine Programme geändert werden müssen. Es ist lediglich die Installation eines TRT-Systems, ein speziell modifizierter

DNS-Server und eine Routing-Konfiguration, welche dafür sorgt, dass die für das TRT-System bestimmten Pakete zu diesem geroutet werden, notwendig.

11.4.7 Vor- und Nachteile der Strategien

Dual-Stack

Wie wir gesehen haben, wird ein Dual-Stack Router vor allem an der Schnittstelle zwischen einem IPv4- und einem IPv6-Netz verwendet. Es wäre nun aber auch möglich, ein ganzes Netzwerk nur mit Dual-Stack Routern auszurüsten. Für eine Firma beispielsweise, die Applikationen mit IPv4 und IPv6 betreibt, wäre das sicherlich ein Vorteil. Die Einrichtung von komplizierten Tunnels würde entfallen. Auf der anderen Seite gäbe es eine Reihe von Nachteilen: Alle Router müssten auf den neusten Stand oder sogar ausgetauscht werden, und bräuchten ausserdem, sowie auch die Hosts, immer zwei Adressen für jedes Interface, eine IPv4- und eine IPv6-Adresse. Das Routing-Protokoll müsste auch parallel betrieben werden, und es müsste genug Speicher vorhanden sein für beide Routing-Tabellen. [30]

Diese Strategie auf das globale Internet anzuwenden wäre natürlich ebenso unvorstellbar, wie wenn das gesamte Internet, Router und Hosts, von einem Tag auf den anderen nur noch IPv6-fähig wäre. Sprich: Ob auf einen Schlag alle Knoten Dual-IP-fähig, oder IPv6-fähig sind, würde keinen Unterschied machen. Deshalb ist diese Strategie schon eher für in sich abgeschlossene Organisationen wie Unternehmen oder ISP gedacht. Man muss damit rechnen, dass es noch für eine lange Zeit Inseln geben wird, zu Beginn sind dies noch IPv6-Inseln, zu einem späteren Zeitpunkt werden es IPv4-Inseln sein.

Tunneling

Wenn jetzt eine Unternehmung IPv6 einführt, und nehmen wir weiter an, dass diese Unternehmung auf verschiedene Standorte verteilt ist, dann wird sie nicht umhin kommen, das Tunneling anzuwenden, denn nur so kann sie auch von den Erweiterungen von IPv6 profitieren. Alle Feauteres von IPv6 können natürlich nicht ausgenutzt werden, so zum Beispiel die Angabe von Quality of Services, denn schliesslich werden die Pakete immer noch über ein IPv4-Netz geroutet, das den QoS keine Beachtung schenkt. Dafür sind nur zwei Dual-Stack-Router notwendig, und allenfalls deren Konfiguration, wenn ein konfigurierter Tunnel verwendet wird. Die Umstellung der Hosts auf IPv6 ist dabei auch nicht zu unterschätzen, genau wie bei der reinen Dual-Stack-Strategie.

Beim Tunneling muss noch hinzugefügt werden, dass immer noch IPv4-Adressen benötigt werden. Wie bereits beschrieben wurde, setzt sich eine Adresse für das Tunneling aus dem Präfix 2002, der IPv4-Adresse und der Interface-ID zusammen. Der Dual-Stack-Router, der dieses Netz mit einem IPv4-Netz verbindet, zeigt nach aussen hin immer noch diese IPv4-Adresse. Das automatische Tunneln ermöglicht es, dass ein Tunnel über weite Strecken aufgebaut werden kann. Adressiert wird hierbei aber immer mit dem Präfix 2002 und der alten IPv4-Adresse. Denn mit einer global vergebenen IPv6-Adresse kann solange nicht adressiert werden, bis der ganze Weg zwischen zwei Hosts vollständig IPv6-fähig ist.

Dies soll zeigen, dass die Routing-Effizienz, die mit IPv6 schliesslich einiges besser sein soll, mit dem Tunneln keinen Unterschied spürt, sondern erst richtig zum Tragen kommt, wenn die IPv4-Inseln nach und nach verschwinden.

Herausforderungen

Das Internet hat bis anhin gut funktioniert, und wo Fehler oder Mängel festgestellt wurden, konnte man sie in kleinen Schritten beheben. Aber ein neues Protokoll einzuführen ist mit Ungewissheit verbunden, auch wenn ausgiebig getestet wurde. Fehler würden sich gravierend auf die Wirtschaft auswirken, vor allem bei Applikationen, die 'mission-critical' sind, wie es [31] ausdrückt.

Netzwerkadministratoren und ISPs wissen nicht, wann und wie auf IPv6 zu migrieren. Sie brauchen die Gewissheit, dass IPv6 einen fliessenden Übergang und Koexistenz mit IPv4 bringen wird, das Netzwerk bei Ausfällen schnell korrigiert werden kann, die Quality of Service sichergestellt ist, und dass die Netzwerksicherheit verbessert wird. Netzwerkhersteller stehen vor der Herausforderung, Router mit Dual-Stack-Technologie herzustellen, wobei die Performance trotz doppeltem IP-Stack nicht beeinträchtigt werden soll. Aber auch Geräte, die Tunneling oder NAT-PT unterstützen, sind ein neues Gebiet für sie.

11.4.8 Kosten der Migration

Es stellt sich natürlich die Frage, weshalb ein funktionierendes Netz umgerüstet werden soll, wenn kein direkter Nutzen ersichtlich ist, bzw. die Ausgaben nicht mindestens durch zusätzlichen Gewinn gedeckt sind. Nach Dittler [32] wird die Einführung von IPv6 schneller vorangehen, wenn es neue Bereiche mit zusätzlichem Wachstum, oder auf IPv6 basierende Anwendungen geben wird. Wenn sich die Notwendigkeit aber nur auf den grösseren Adressraum stützt, wird das Internet noch etliche Jahre mit IPv4 auskommen können.

Obwohl die Umstellung noch nicht in Aussicht ist, sind aber schon viele Software- und Hardwarehersteller auf IPv6 eingestellt, und haben das neue Protokoll schon integriert. So unterstützt zum Beispiel Windows XP, Mac OS X oder Sun Solaris 8, und auch die meisten UNIX-Versionen das neue IP. Etliche Hosts sind dabei in dieser Aufzählung nicht enthalten. Bei den Herstellern von Routern sind unter anderen Cisco, Hitachi, 6WIND und Extreme Networks, die IPv6 standardmässig einbauen. [33] Wo aber zusätzliche Kosten anfallen könnten, wäre bei Firewall-Herstellern, die für ein Upgrade etwas verlangen könnten.

Die Kosten für die Einführung von IPv6 fallen aber nicht nur beim Kauf von Geräten und Software an: Mitarbeiter müssen geschult werden, die Geräte richtig konfiguriert werden sowie die laufenden Betriebskosten. Nach Dittler [32] soll die Konfiguration von kleineren Netzen unter IPv6 aber deutlich einfacher sein, weil die Endgeräte sich automatisch konfigurieren, und ein Verwalter könnte praktisch entfallen. Er spricht ausserdem noch den Vergleich zu NAT bei heutigen Systemen mit IPv4 an: "‘...Außerdem ist die Verwaltung

von Netzen mit festen Adressen deutlich einfacher und das Suchen von Fehlern in Netzen mit dynamisch zugeteilten Adressen ist für den Verwalter ein wahrer Alptraum."'

Ausser Frage steht, dass zu Beginn sicher hohe Umstellungskosten anfallen werden, aber auch operationelle Kosten im kurz- bis mittelfristigen Bereich werden höher sein als bisher. Und zwar so lange, wie IPv4 und IPv6 miteinander koexistieren werden. [34]

11.4.9 Bereits existierende IPv6-Netze - Zwei Beispiele

Dolphins Network Systems AG, Zürich

Dolphins Network Systems AG ist ein Internet Service Provider mit Sitz in Zürich und bietet verschiedene Dienstleistungen wie Webserver oder Mailserver an. Ihre Kunden bedient sie über ein ADSL-Netz und unterhält einen eigenen Backbone. Die Netzwerk Topologie von Dolphin ist aus der Grafik 11.14 ersichtlich. Sie unterhält 'Points of Presence' in Zürich, Regensdorf, Glattbrugg, Buchs und Otelfingen, und die Linien dazwischen stellen Backbone-Verbindungen dar. Der gesamte Backbone ist bereits unter IPv6 in Betrieb. Auch einige Server sind sowohl mit IPv4 als auch mit IPv6 erreichbar.

Wie aus den News der Dolphins Webseite hervorgeht, hatten sie bereits im Jahre 2002 verschiedene Router auf Dual-Stack upgraded. Im September 2002 erhielten sie dann eine globale Unicast-Adresse von RIPE (Réseaux IP Européens), und kurz darauf war der ganze Backbone Dual-Stacked und über Peerings mit Tiscali und Cyberlink verbunden. Im Jahre 2003 stellten sie ihren ADSL-Kunden IPv6 zur Verfügung, und waren via Swisscom mit Euro6IX (European IPv6 Internet Exchanges Backbone) verbunden. [35]

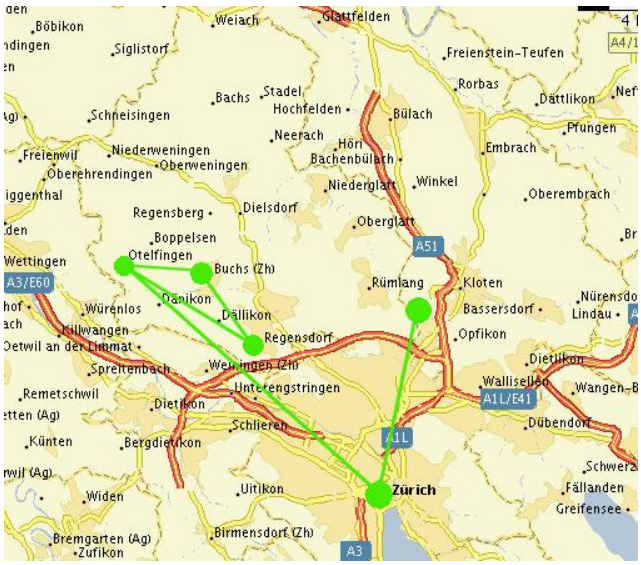

Figure 11.14: Dolphins Netzwerk Topologie Der Verkehr wird über zwei Carriers und diversen Peerings ins Internet geleitet.

SKY Perfect Communications Inc., Japan

SKY Perfect Communications ist der grösste Anbieter von Fernseh- und Kommunikationsübertragungen per Satellit in Japan [36]. Zu Beginn war ihre Infrastruktur so aufgebaut, dass es nur möglich war, Fernsehsendungen via Satellit zu übertragen. Die Kunden hatten dafür eine Set-Top-Box vor den Fernseher geschaltet. Das Problem war nun, dass an vielen Standorten keine Möglichkeit bestand, eine Satellitenschüssel aufzustellen, beispielsweise Hauseigentümer, die dies ihren Mietern verboten. Die Lösung war, die Sendungen über das Internet mittels Multicast an die Haushalte zu vertreiben. Da dies mit IPv4 aber nicht möglich war, kam man schnell auf die Idee, dass nur IPv6 Abhilfe schaffen konnte.

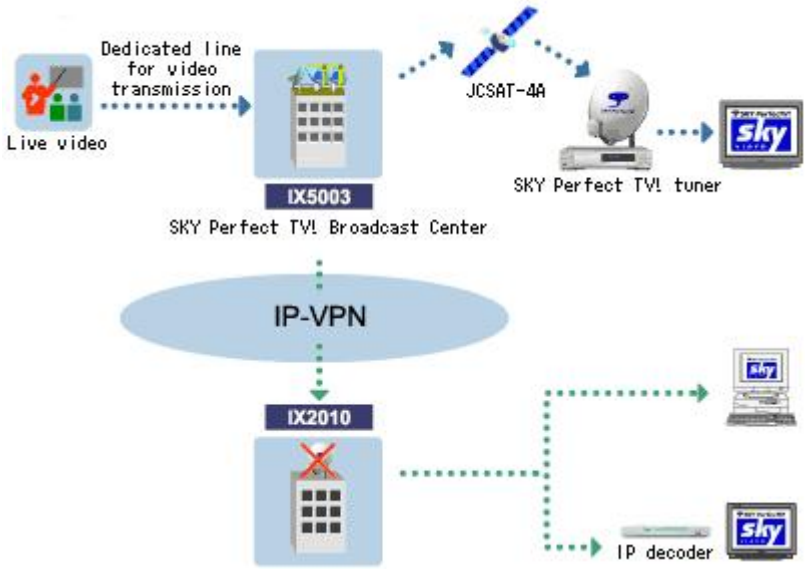

Figure 11.15: SKY Stream: Übertragung von Video-Daten

Auf der Grafik 11.15 ist ersichtlich, dass es auch möglich ist, eigene Videos über Multicast zu übertragen, zum Beispiel eine Live-Übertragung einer Vorlesung. Diese Daten werden dann zuerst an SKY Perfect geschickt. Eine private Übertragung würde dann nur über das Internet, und nicht über Satellit vertrieben werden. Dabei wandern die Video-Daten zuerst über einen VPN-Tunnel, zur Zeit noch über IPv4, an eine Verteilerstelle, von wo aus dann das Multicast per IPv6 gestartet wird.

11.5 Fazit

Es sind genügend Vorbereitungen getroffen worden, um einen reibungslosen Übergang von IPv4 auf IPv6, und vor allem deren Koexistenz zu gewährleisten. Es wird auch immer weiter aktiv getestet, ganze Umgebungen mit Routern und Hosts werden aufgestellt, um die Tests so realistisch wie möglich zu machen. Trotzdem ist man nie davor gefeit, dass die Realität wieder ein wenig anders sein könnte. Verschiedene Netzwerke und Backbones laufen schon heute mit IPv6. Doch bis einmal eine kritische Masse erreicht ist, wird sich

das Neue nur schleppend durchsetzen. Es braucht auch genügend Anreize in Form von neuen IPv6-basierten Applikationen, damit die Anwender den Nutzen des neuen IP sehen. Vielleicht kommt der Durchbruch aber auch von Asien her. Es wäre nicht weit hergeholt, zu sagen, dass in naher Zukunft der grösste Teil Asiens nur noch IPv6-Netze betreibt. Dann steht die Wirtschaft in Europa und Nordamerika unter Druck. Vielleicht gibt man sich dann zufrieden, dass die Adressen mit SIIT und NAT-PT übersetzt werden können, oder aber es stellt sich heraus, dass die Performance doch nicht reicht.

Irgendwann wird sich IPv6 doch durchsetzen, ob man nun will oder nicht, denn die neuen Netzwerkgeräte sind dafür ausgelegt, und es ist nur eine Frage der Zeit, bis die IPv4-Geräte altershalber ersetzt werden müssen.

Bibliography

[1] Douglas E. Comer: Computernetzwerke und Internets, Pearson Studium, 2002

[2] V. Fuller, T. Li, J. Yu, K. Varadhan: Classless Inter-Domain Routing (CIDR): an Address Assignment and Aggregation Strategy, RFC 1519, September 1993

[3] Stefan Strobel, Firewalls und IT-Sicherheit, dpunkt.verlag GmbH, 2003

[4] Information Sciences Institute University of Southern California: Internet Protocol, RFC 791, September 1981

[5] Y. Rekhter, B. Moskowitz, D. Karrenberg, G. j. de Groot, E. Lear: Address Allocation for Private Internets, RFC 1918, Februar 1996

[6] 6WIND, http://www.6wind.com.frame.php, 20.1.2005

[7] S. Bradner und A. Mankin: Next Generation (IPng) White Paper Solicitation, RFC 1550, Dezember 1993

[8] Deering S.E.: SIP - Simple Internet Protocol. IEEE Network Magazine, Band 7, S. 16-28, Mai/Juni 1993

[9] Francis P.: A Near-Termn Architecture for Deploying Pip. IEEE Network Magazine, Band 7, S. 30-37, Mai/Juni 1993

[10] S. Deering und R. Hinden: Internet Protocol Version 6 (IPv6), RFC 2460, Dezember 1998.

[11] S. Deering und R. Hinden: Internet Protocol Version 6 (IPv6), RFC 1883, Dezember 1995

[12] K. Nichols, S. Blake, F. Baker und D. Black: Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers, RFC 2474, Dezember 1998

[13] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang und W. Weiss: An Architecture for Differentiated Service, RFC 2475, Dezember 1998

[14] H. Wiese: Das neue Internetprotokoll IPv6, Hanser Verlag, 2002

[15] J. Moy: OSPF Version 2, RFC 2328, April 1998

[16] R. Coltun, D. Ferguson und J. Moy: ISPF for IPv6, RFC 2740, Dezember 1999

[17] R. Sedgewick: Algorithmen, Addison-Wesley, 1992

[18] C. Eckert: IT-Sicherheit, Oldenburg Verlag 2. Auflage, 2003

[19] S. Thomson und T. Narten: IPv6 Stateless Address Autoconfiguration, RFC 2462, Dezember 1998

[20] Das deutsche IPv6 Forum, http://www.ipv6-net.de/themen/uebe/, Download am 30.12.2004

[21] Preserving Digital Information. www.rlg.org/ArchTF/tfadi.randr.htm

[22] R. Gilligan: Transitions Mechanismen for IPv6-Hosts and Routers, RFC1933) April 1996

[23] S. Thomson: DNS Extensions to support IP version 6, RFC 1886, December 1995

[24] R. Gilligan: Transitions Mechanismen for IPv6-Hosts and Routers, RFC2893, August 2000

[25] Transition Mechanisms and Packet Tunneling, Heiko Raue, 1997

[26] B. Carpenter: Connection of IPv6 Domains via IPv4 Clouds, RFC3056, February 2001

[27] E. Nordmark: Stateless IP/ICMP Translation (SIIT), RFC2766, February 2000

[28] G. Tsirtsis: Network Address Translation - Protocol Translation (NAT-PT), RFC2766, February 2000

[29] J. Hagino: IPv6-to-IPv4-Transport-Relay-Translator, RFC3142, June 2001

[30] iX, Magazin für Profesionelle Informationstechnik, Februar 2005

[31] Dean Lee, Elliott Stewart: Internet Protocol version 6 (IPv6) Conformance and Performance Testing, Ixia, 27.1.2005

[32] Stefan Krempl: Generalüberholung für das Internet, Round-Table-Gespräch (c't 20/1999), mit Hans Peter Dittler, Jürgen Rauschenbach

[33] Host-, Router-Implementations, http://playground.sun.com/pub/ipng/html/ipng-implementations.html

[34] Ken Wieland: Addressing the IPv6 issue, www.telecommagazine.com, Mai 2002

[35] Dolphins Network Systems AG, Zürich, Full Service Provider

[36] SKY Perfect Communications Inc., http://www.skyperfectv.co.jp/skycom/e/