University of Zurich
Department of Informatics

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

# Mobile Systems I

TECHNICAL REPORT – No. ifi-2005.06

July 2005

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. ifi-2005.06, July 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 works on research and teaching in the area of communications. One of the driving topics in applying communications technology is addressing investigations on mobility aspects and support for mobile users. Therefore, during the summer term SS 2005 a new instance of the Mobile Systems seminar has been prepared and students as well as supervisors worked on this topic.

Even today, the increasing number of mobile and wireless networks as well as their users or customers drive many developments of systems and protocols for mobile systems. The areas of underlying networking and development technology, of services assisting security or Quality-of-Service (QoS), and of mobility support determine an important part of future wireless networks. Therefore, this year's seminar addressed such areas in more depth. The understanding and clear identification of problems in technical and organizational terms have been prepared and challenges as well as weaknesses of existing approaches have been addressed in a mobile and wireless environment. All talks in this seminar provide a systematic approach to judge dedicated pieces of systems or proposals and their suitability.

## Content

This new and first edition of the seminar entitled "Mobile Systems I" discusses a number of selected topics in the area of mobile communication. The first talk "Integrierte Architektur für UMTS und WLAN" summarizes and evaluates different approaches for the integration of UMTS and WLAN networks in order to consolidate the advantages of both technologies. Talk two "DVB-basierte Rundfunksysteme" provides an overview of digital broadcasting systems for audio and video data. It also discusses the IP-based data transfer over broadcast networks. "Technology and Use of RFID" as talk three gives an introduction to the RFID technology and its various application fields. The fourth talk addresses "Fast Handover in Mobile IPv4", which aims to provide a better handover support in mobile IP scenarios. Talk five "The Session Initiation Protocol in Mobile Environments" presents the SIP protocol used to setup and negotiate media sessions. The talk also discusses related protocols and shows up issues in case of mobile environments. The sixth talk "Verfahren zur Umsetzung von mobilen DRM-Systemen" addresses the aspects of digital right management systems.

Talk seven “Mobile Spiele - Ein Überblick über verschiedene Technologien” outlines mobile gaming and its economic potentials. The talk presents a check list of criteria to provide means for an evaluation. “Security in WLAN” as talk eight addresses security mechanisms in WLAN, including WEP, WPA. It discusses several security problems which arise in wireless networks and presents possible countermeasures. After the security aspects in WLAN, the ninth talk “Security in GSM and GPRS Networks” addresses the same problems in GSM and GPRS networks. Talk ten “Diameter in Wireless Environments” presents the AAA architecture and the Diameter protocol, considered as the successor of RADIUS. The talk also discusses AAA support for mobile IP based on Diameter. Talk eleven continues with “Ansätze zur Umsetzung mobiler Ticketing-Systeme” and outlines approaches for mobile ticketing systems. Finally, talk twelve “Technologies Beyond 3G” gives and introduction to next generation mobile network technologies.

# 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 Mobile Systems, 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, July 2005*

# Contents

# Kapitel 1

# Integrierte Architektur für UMTS und WLAN

*Thomas Loher, Nikola Mijatovic*

*Die neue Mobilfunktechnologie UMTS bietet dem Benutzer nicht nur die Übertragung von Sprache, sondern auch von Daten mit zumutbarer Bandbreite. Ausserdem unterstützt sie dessen Mobilität durch eine flächendeckende Versorgung und durch einen einheitlichen internationalen Standard. Eine andere Möglichkeit der drahtlosen Datenübertragung ist WLAN. Diese Technologie bietet ihren Benutzern höhere Übertragungsraten als UMTS, allerdings sind die Abdeckung und die Mobilität deutlich eingeschränkt. Deshalb wird versucht, UMTS und WLAN in einem Netzwerk nahtlos zu vereinen und damit die jeweiligen Vorteile zu nutzen. Es gibt grundsätzlich zwei Vorschläge, wie eine solche integrierte Architektur aussehen kann. Bei der engen Kopplung wird versucht, die UMTS-Infrastruktur in den Bereichen, wo sie sich bewährt hat, zu übernehmen. Demgegenüber verfolgt die lose Kopplung das Ziel, möglichst wenige Änderungen an der bestehenden WLAN-Struktur vornehmen zu müssen. Da die lose Kopplung im allgemeinen eher favorisiert wird, werden wir v.a. auf diese eingehen und diskutieren, wie der Datenaustausch mit einem externen Netz ablaufen soll und wie sich daraus ein Vorgehen für die Abrechnung der genutzten Dienste ergibt. Ausserdem gehen wir ein auf die Authentifikation und die Autorisierung, die mittels einer zusätzlichen Komponente realisiert wird. Schliesslich behandeln wir die Sicherheit und die Service-Qualität in der losen Kopplung.*

## Inhaltsverzeichnis

## 1.1 Einleitung

In der heutigen Gesellschaft besteht ein wachsendes Bedürfnis nach der Übertragung von Daten bei hoher Mobilität des Benutzers. So sind Geschäftsreisende zunehmend daran interessiert oder sogar darauf angewiesen, auch unterwegs E-Mails zu lesen, Informationen aus dem Internet abzurufen oder umfangreiche Dateien wie Multimedia-Dokumente zu empfangen bzw. abzusenden. Aber auch bei den Privatkonsumenten steigen die Ansprüche immer weiter an, beispielsweise beim Video-Streaming.

In den letzten paar Jahren sind drahtlose lokale Netze (WLANs) stark in Mode gekommen. So ist auch an der Universität Zürich seit etwa zwei Jahren ein solches im Betrieb. Die WLANs bieten dem Benutzer bereits hohe Übertragungsraten von bis 54 Mbit/s, falls er sich in der Nähe eines Zugangspunktes (AP) aufhält, d. h. in einem Radius von etwa 100 m im Freien und etwa 30 m in Gebäuden um den AP [6]. Typischerweise befinden sich diese APs an vielbesuchten Plätzen wie Bahnhöfen, Hotels, Einkaufszentren oder Bürogebäuden. Ein WLAN ist relativ billig einzurichten und zu unterhalten. Allerdings gibt es auch Probleme: der beschränkte Abdeckungsbereich und vor allem die Sicherheit. Der bisher verwendete Verschlüsselungsalgorithmus ist geknackt worden. Wir werden im Abschnitt 1.2 einige Grundlagen über WLAN bereitstellen.

Die Mobilfunk-Technologien der dritten Generation (3G) wie UMTS bieten die Übertragung von Sprache und auch Daten mit nahezu globaler Abdeckung. Dazu werden Zellen unterschiedlicher Grösse verwendet (Makro-, Mikro- und Pikozellen), wobei in den kleinsten die Datenrate 2 Mbit/s betragen soll. Dieser Wert ist allerdings von mehreren Bedingungen abhängig und eher als theoretische obere Grenze zu verstehen. Andererseits wird die Mobilität des Benutzers im 3G-Netz durch ein gut funktionierendes Handover und Roaming unterstützt. Ein generelles Problem der 3G-Technologie stellen aber die sehr hohen Kosten für den Aufbau des Netzes dar. Da die Funkfrequenzen im lizenzierten Bereich liegen, mussten die potenziellen Anbieter eine Lizenz ersteigern. In der Schweiz hat die Eidgenössische Kommunikationskommission (ComCom) vier UMTS-Konzessionen vergeben [7]. Swisscom, Sunrise und Team 3G bezahlten dafür je 50 Millionen Franken und Orange sogar 55 Millionen Franken. In Deutschland lagen die entsprechenden Beträge noch um ein Vielfaches höher. Näheres zum Thema UMTS findet man in Abschnitt 1.3.

Wenn man die Eigenschaften der beiden Technologien WLAN und UMTS vergleicht, stellt man fest, dass sich die Vor- und Nachteile komplementär verhalten. Wenn eine der Technologien in einem Bereich einen Nachteil aufweist, hat die andere im selben Bereich einen Vorteil. Dies legt nahe, die WLAN- und die UMTS-Technologie in einem gemeinsamen Netzwerk zu verbinden, um so die Vorteile beider Technologien zu nutzen. Es gibt verschiedene Ansätze, wie diese Verbindung realisiert werden könnte. Wir werden auf die beiden Hauptvorschläge eingehen, die sogenannte *enge* und die *lose Kopplung*.

Die enge Kopplung (siehe Abschnitt 1.4.2) geht davon aus, dass sich wesentliche Bereiche der Infrastruktur im 2G- und damit auch im 3G-Netz bewährt haben und deswegen auch in einem integrierten Netz wieder verwendet werden sollten. Dazu zählen etwa das Mobilitäts-Management, die Sicherheit oder die Verwaltung der Abonnenten. Wie die Bezeichnung schon andeutet, ist die Beziehung zwischen WLAN und dem 3G-Netz eng, und das bringt auch einige Probleme mit sich, wie im betreffenden Abschnitt ausgeführt wird.

Bei der losen Kopplung sollen das WLAN und das 3G-Netz möglichst eigenständig bleiben. Diese Form der integrierten Architektur ist flexibler und technisch sowie finanziell einfacher realisierbar, weswegen sie im allgemeinen bevorzugt wird. Andererseits erfordert die lose Kopplung gewisse Änderungen an der Struktur der bestehenden Mobilfunknetze. Der Vorschlag [1] (siehe auch [4]) sieht vor, dass einige neue Komponenten im 3G-Netz installiert werden. Darauf wird in den Abschnitten 1.5.2 und 1.5.4 eingegangen.

Ein wichtiges Thema ist auch die Sicherheit. Wir werden sehen, wie (im Fall der losen Kopplung) das bewährte UMTS-Verfahren für die Authentifikation und die Autorisierung auch weiter verwendet werden kann (Abschnitt 1.5.2). In Abschnitt 1.5.3 das Problem, wie in der integrierten Architektur die Dienstqualität sichergestellt wird, und schliesslich befassen wir uns in 1.5.4 mit der Abrechnung im integrierten Netz.

## 1.2 WLAN

WLAN steht für Wireless LAN und ist der Überbegriff für die Technologien der drahtlosen Vernetzung und Kommunikation zwischen Rechnern.

Es sind einige Technologien auf dem Markt, die als WLAN agieren:

- IEEE 802.11 Standard, was als die meistverbreitete WLAN Technologie gilt, und darum auch öfters selbst als WLAN bezeichnet wird.
- HIPERLAN/2
- HiSWAN (Japan)
- Wireless ATM
- HomeRF

In folgendem soll genauer auf die erste Technologie eingegangen werden, nämlich den 802.11 Standard.

Der 802.11 Standard ist ein Kind der 802 Netzwerksspezifikationsfamilie. Er definiert eine Protokollarchitektur für drahtlose Kommunikation. Das Gemeinsame an allen 802 Standards ist die oberste Schicht des Protokollstapels (siehe Abbildung 1.1), nämlich die Logical Link Control Schicht. Sie ist allen gemeinsam aus dem Grund, dass die darüber liegende dritte Schicht des OSI Layer Modells auf die gleiche Schnittstelle zugreifen kann, unabhängig von der Technologien für die zweite und erste Schicht [10].

Der 802.11 Standard kennt zwei Betriebsmodi: den Infrastruktur Modus und den Ad-hoc-Modus. Der Infrastruktur Modus ist der bekanntere der beiden. Dabei wird den Benutzern eine Infrastruktur als Teil des gesamten Netzwerks zu Verfügung gestellt, an welcher sie ihre Rechner über Wireless Schnittstellen an das Netzwerk anbinden können. Die Schnittstellen des Netzwerks heissen Access Points (AP). Sie sind nicht nur für das Anbinden der

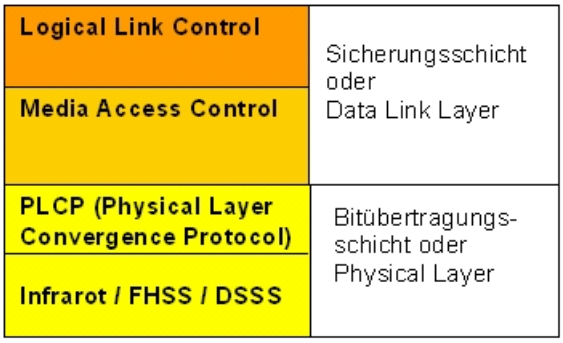


Abbildung 1.1: 802.11 Protokollstapel

Clients an das grundlegende LAN verantwortlich sondern bieten auch gewisse Dienste an wie Synchronisation der Uhren oder Power Management.

Der zweite Modus, der Ad-hoc-Modus, bezeichnet die Verbindung der mobilen Clients untereinander. Die Verbindung ist direkt und ohne jegliche Management Instanzen. Anzumerken ist noch, dass keine Verbindung zu einem drahtlosen LAN existiert, sprich die mobilen Clients bilden einen geschlossenen WLAN für sich.

Nun kommen wir zu den Vor- und Nachteilen von WLAN bzw. 802.11 Standard. Zu den wohl grössten Vorteilen gehört die hohe Übertragungsrate beim mobilen Einsatz. Die maximale Rate beträgt 54 Mbit/s. Des Weiteren greift der 802.11 Standard auf die lizenzfreien Frequenzen. Die Hardware hat sich soweit entwickelt, dass sie sogar in privaten Haushalten eingesetzt wird, d.h. die Verfügbarkeit ist sehr gross. Stichwort SOHO (Small Office Home Office). Soviel zu den Kostenaspekt. Nicht zu vergessen ist auch der Kabelsalat der üblichen LANs, der mit WLANs wegfällt.

Von den Nachteilen ist wohl die Sicherheit der grösste. Jede Funkverbindung ist abhörbar. Auch wenn sie verschlüsselt ist, kann man sie trotzdem abfangen. Die Verschlüsselungen haben sich auch nicht als zuverlässig erwiesen. Zu diesem Thema werden wir etwas später zurückkehren. Ein weiterer technischer Nachteil sind die Roaming- und Mobilitäts- Mechanismen. In der Mobilität müssen die Benutzer einen Kompromiss eingehen mit den Datenraten. Im WLAN-Jargon wird das Roaming als der Handover des mobilen Clients von einem AP zu einem anderen verstanden. Das Handover der Mobilen Stationen (MS) von einem Access Point zum anderem läuft folgendermassen ab:

1. Die MS erkennt die Abnahme der Verbindungsstärke zu einem AP.
2. Sie versucht danach einen anderen AP zu finden.
3. Ist ein AP gefunden worden, so registriert sie sich dort und wartet auf die Quittung.
4. Der neue AP informiert den Rest des Netzwerkes über den Wechsel und quittiert der MS den Zugang.

Das Informieren anderer AP ist an ein gemeinsames Netzwerk angewiesen, und genau dies stellt ein Problem dar. Heutzutage ist die Mobilität möglich nur bei einem Provider und

nicht bei mehreren gleichzeitig, d.h. man kann sich nicht nahtlos in Netzen bewegen, die von verschiedenen Providern verwaltet werden.

Ein Stichwort das des Öfteren im Zusammenhang mit WLAN anfällt ist Hotspot. Hotspots sind WLAN AP an öffentlichen Plätzen so wie Flughafen, Hotels, Business Zentren, etc. Ziel ist es einem öffentlichen Zugang zu Internet zu Verfügung zu stellen und somit die Mobilität der Benutzer zu erhöhen. Natürlich trägt es zum Image bzw. erweitert es die Dienstleistung.

## 1.3 UMTS

### 1.3.1 Mobilfunk-Generationen

In den achtziger Jahren wurden in Europa, den USA und in Japan Mobilfunksysteme der ersten Generation (1G) eingeführt. Der einzige Zweck dieser noch analogen Systeme war die Übertragung von Sprache, und sie waren untereinander meistens nicht kompatibel.

Rund zehn Jahre später wurde die erste Generation durch die zweite abgelöst, bei der neben der Sprache beschränkt auch Daten übertragen werden können. In der Schweiz wurde GSM im Jahr 1993 kommerziell in Betrieb genommen, aber es gibt auch bei den 2G-Systemen mehrere inkompatible Standards. So werden etwa in Asien oder Amerika andere Systeme verwendet. Das kann für Reisende ein Problem sein — in einem Land, in dem das eigene Netz nicht verfügbar ist, kann der Benutzer sein Mobiltelefon eventuell gar nicht gebrauchen. Mit der Zeit wurden mehrere Erweiterungen von GSM eingeführt, etwa GPRS, die deutlich höhere Übertragungsraten ermöglichten.

Das alles reichte aber noch nicht aus, um auch grosse Mengen von Daten in einer zumutbaren Zeit übertragen zu können. Daher wurde in den neunziger Jahren die Planung der dritten Generation begonnen. Einer der Vorschläge, die sich daraus ergaben, war UMTS (Universal Mobile Telecommunications System), das im Rahmen des 3GPP (3rd Generation Partnership Project) entwickelt wurde. Die zugrundeliegenden Ziele waren die folgenden (nach [9]):

- Unterstützung von Multimedia: Die Endgeräte sollten in der Lage sein, gleichzeitig Dienste unterschiedlicher Art zu empfangen, etwa Sprache, Bildtelefon oder ganz allgemein Daten. Dies sollte dank einer Übertragungskapazität von bis zu 2 MBit/s möglich werden. Ausserdem sollte das neue Netz fähig sein, Dienste dieser Art auch anzubieten.

- Es sollte anders als bei den 2G-Systemen nun eine einheitliche Norm geschaffen werden. Damit zusammenhängend wollte man auch ein weltweites Roaming ermöglichen. Unter Roaming versteht man beim Mobilfunk die Nutzung eines fremden Netzes, etwa beim Telefonieren im Ausland.

Tabelle 1.1: UMTS-Zellen

| | Durchmesser | Datenrate | Geschwindigkeit |
|---|---|---|---|
| Makrozelle | einige km | 144 Kbit/s | bis 500 km/h |
| Mikrozelle | bis ca. 2 km | 384 Kbit/s | bis 120 km/h |
| Pikozelle | einige m | 2 Mbit/s | bis 10 km/h |

- Das neue System sollte abwärtskompatibel zu den 2G-Systemen sein. Das bedeutet einerseits, dass Dienste, die im 2G-Netz angeboten werden, weiterhin verfügbar sein sollten (beispielsweise SMS). Andererseits heisst das Transparenz des Netzes. UMTS wird (mindestens am Anfang) nicht flächendeckend verfügbar sein, sondern die UMTS-Bereiche werden Inseln im GSM-Gebiet bilden. Verlässt nun beispielsweise ein Anwender den UMTS-Bereich, so sollen die gerade benutzten Dienste von GSM weiterversorgt werden und nicht einfach abrupt abbrechen.

Die Abwärtskompatibilität ist auch aus Kostengründen wünschenswert. Die Betreiber der 2G-Netze haben sehr hohe Geldbeträge in den Aufbau ihrer Netze investiert, und man rechnet etwa mit 5 bis 10 Jahren Betriebsdauer bis zur Amortisation dieser Investitionen.

Im Gegensatz zu den GSM-Zellen gibt es bei den UMTS-Zellen verschiedene Grössen (siehe Abbildung 1.2).

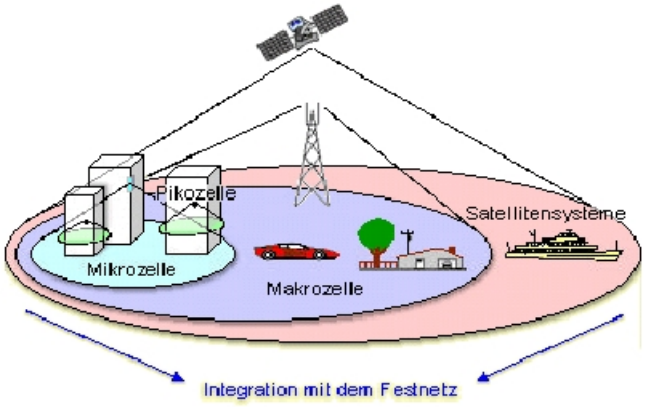


Abbildung 1.2: UMTS-Zellen

Die kleinsten sind die Pikozellen, die zur Versorgung in vielbesuchten Gebäuden wie Einkaufszentren oder Bahnhofhallen vorgesehen sind. Die Mikrozellen sind für Gebiete innerhalb von Städten vorgesehen, und Makrozellen sollen dünn besiedelte Gebiete abdecken. Schliesslich werden, um eine globale Abdeckung zu erreichen, noch Satellitenzellen benötigt. Die Übertragungkapazität ist sowohl von der Zelle als auch von der Geschwindigkeit abhängig, mit der sich der Benutzer fortbewegt. Die Details zu den Zellhierarchien sind in der Tabelle 1.1 enthalten.

Für die Anmeldung im UMTS-Netz wird das Mobiltelefon mit der USIM-Karte ausgerüstet. Diese ist eine weiterentwickelte SIM-Karte und enthält die nötigen Algorithmen und Informationen für die Authentifikation und die Vereinbarung der kryptographischen

Schlüssel (AKA, Authentication and Key Agreement). Darauf wird im Abschnitt 1.5.2 detaillierter eingegangen.

### 1.3.2 UMTS-Komponenten

Wir wollen nun ganz kurz auf diejenigen Komponenten im UMTS-Netz eingehen, die wir zur Beschreibung der integrierten Architektur brauchen werden. Diese sind in der Abbildung 1.3 dargestellt. Für eine umfassende Darstellung sei etwa auf [9] verwiesen.

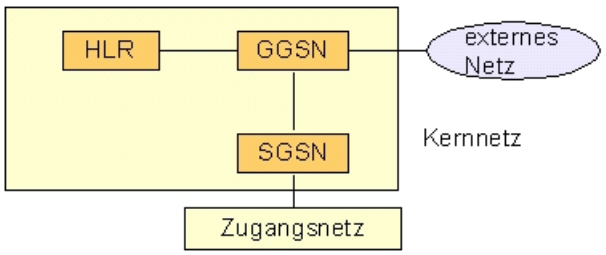


Abbildung 1.3: Komponenten des UMTS-Netzes

Das UMTS-Netz besteht aus dem Zugangs- und dem Kernnetz. Das Zugangsnetz enthält diejenigen Komponenten, die für die Übertragung der Funksignale und die Verbindung zum Kernnetz zuständig sind. Das Kernnetz basiert im wesentlichen auf dem GSM- und GPRS-Netz. Es besteht selbst aus drei Teilen, der leitungs- und der paketvermittelten Domäne sowie einem dritten Bereich, der von beiden genutzt wird. Die erste ist für die Sprachübertragung zuständig und wird hier nicht weiter beachtet. Die folgenden beiden Komponenten befinden sich im paketvermittelten Bereich:

Ein *SGSN* (Serving GPRS Support Node) ist jeweils für ein bestimmtes geographisches Gebiet zuständig. Er vermittelt Daten aus dem Kernnetz an die Benutzer in seinem Gebiet und umgekehrt. Ausserdem verwaltet er die Sitzungen und sorgt für die Mobilität. Wenn der Benutzer das Gebiet seines SGSN verlässt, so übergibt dieser die Verbindung an den neu zuständigen SGSN.

Ein *GGSN* (Gateway GPRS Support Node) verbindet das Kernnetz mit einem oder mehreren externen IP-Netzen, z. B. dem Internet. Er leitet die Daten aus dem externen Netz an den zuständigen SGSN weiter.

Der *HLR* (Home Location Register) befindet sich im gemeinsam genutzten Bereich. Er enthält eine Datenbank mit Informationen über die Abonnenten. So wird etwa beim Einschalten des Mobiltelefons aus dem Zugangsnetz eine Verbindung zum entsprechenden SGSN hergestellt, der dann die zur Anmeldung nötigen Informationen aus dem HLR bezieht.

Ein externes Netz ist irgendein IP-Netz, wie etwa das Internet, ein IP-Netz des UMTS-Betreibers selbst oder ein Intranet einer anderen Firma.

Tabelle 1.2: Vergleich zwischen UMTS und WLAN

| | WLAN | UMTS |
|---|---|---|
| Geographische Abdeckung | gering | gut |
| Fähigkeiten zum Handover/Roaming | gering | gut |
| Sicherheit der Datenübertragung | gering | hoch |
| Datenübertragungsrate | hoch | gering |
| Aufbau- und Betriebskosten | gering | hoch |
| Verwaltung der Abonnenten | eher schlechter | gut |

## 1.4 Integrations-Architektur

### 1.4.1 Problematik und Anforderungen

Wir haben uns in den beiden vorhergehenden Kapiteln mit den beiden Technologien WLAN und UMTS beschäftigt und gesehen, dass beide ihre Stärken und Schwächen haben. In Tabelle 1.2 sind diese zusammenfassend dargestellt.

Wir sehen, dass sich die Eigenschaften komplementär verhalten — wenn eine Technologie einen Vorteil hat, so hat die andere beim selben Kriterium ein Problem. Daraus ergibt sich die Frage, ob es nicht möglich ist, UMTS und WLAN in *einem* Netz zu vereinen und dabei die jeweiligen Vorteile zur Kompensation der Nachteile zu nutzen. Es gibt mehrere Vorschläge, was genau die Anforderungen an ein solches integriertes Netz sein sollen und wie ein Standard aussehen könnte.

Das 3GPP hat 6 aufeinander aufbauende Szenarien enwickelt [2] (siehe auch [12] oder [13]), die Anforderungen an eine solche integrierte Architektur beschreiben:

*Szenario 1: gemeinsame Abrechnung und Kundendienst.* Das ist die einfachste Form der Zusammenarbeit, bei der es lediglich um die gemeinsame Abrechnung und Kundenbetreung geht. Darüber hinaus gibt es hier aber keine Zusammenarbeit zwischen dem WLAN und dem 3G-Netz, und deswegen braucht es bei diesem Szenario auch keine Standards.

*Szenario 2: Zugangskontrolle und Abrechnung für beide Netze über das 3G-Netz.* Bei diesem Szenario werden die Authentifikation, Autorisierung und die Abrechnung für das WLAN mit den entsprechenden Verfahren des 3G-Netzes abgewickelt. So kann der WLAN-Benutzer beispielsweise seine SIM-Karte für die Anmeldung benützen, wie er es im 3G-Netz machen würde. Ausserdem wird die Autorisierung vom 3G-Netz aufgrund der Benutzerdaten besorgt. Die 3G-Abonnenten können ihrerseits über das WLAN auf IP-Netze, etwa das Internet, zugreifen. In diesem Szenario werden keine Bedingungen an die Dienste des WLANs gestellt.

*Szenario 3: Zugang zu 3G-Diensten.* Hier ermöglicht es der 3G-Betreiber den WLAN-Abonnenten, aus dem WLAN heraus auch 3G-Dienste zu beanspruchen. Wenn er etwa einen WAP-Gateway unterhält, um seinen Abonnenten WAP-Dienste anzubieten, so sollen diese auch von WLAN-Abonnenten genutzt werden können. Zu beachten ist, dass bei

diesem Szenario für Dienste, die in beiden Netzen benutzt werden können, keine unterbrechungsfreie Fortsetzung über die Grenzen beider Netze hinweg verlangt wird.

*Szenario 4: Kontinuierliche Dienste.* Das Ziel dieses Szenarios ist es, wie beim Szenario 3 den Zugang zu den 3G-Diensten zu ermöglichen und zusätzlich die Fortsetzung dieser Dienste über die Grenzen der Netze zu garantieren. Wenn etwa ein Benutzer eine WAP-Verbindung aufnimmt, so soll diese Verbindung erhalten bleiben, wenn von einem Netz ins andere wechselt. Allerdings wird in diesem Szenario nicht verlangt, dass wirklich alle Dienste so erhalten bleiben. Es kann sein, dass einige Dienste unterbrochen werden bei einem Wechsel des Netzes, etwa wegen der verschiedenen Übertragungskapazitäten oder unterschiedlichen Eigenschaften der Zugangs-Technologie. Ein typisches Beispiel dafür ist ein 3G-Dienst, der nur eine sehr kleine Verzögerung verkraften kann, die vom WLAN nicht erbracht werden kann. In diesem Fall würde der Dienst wahrscheinlich abgebrochen, wenn sich der Benutzer in ein WLAN-Gebiet begibt. Ausserdem wird im Szenario 4 zugelassen, dass sich die Dienstqualität ändert beim Übergang zwischen den beiden Netzen.

*Szenario 5: Nahtlose Dienste.* Dieses Szenario geht einen Schritt weiter als das vierte. Hier geht es um eine nahtlose Übergabe der Dienste zwischen dem WLAN und dem 3G-Netz. Das bedeutet, dass 3G-Dienste über die Netz-Grenzen hinweg nahtlos benutzt werden können, ohne dass der Benutzer einen wesentlichen Unterschied bemerkt.

*Szenario 6: Zugang zu leitungsvermittelten Diensten des 3G-Netzes.* Das Ziel dieses Szenarios ist es, dass der 3G-Betreiber den WLAN-Abonnenten Zugang zu den leitungsvermittelten Diensten (etwa normale Sprachübertragung) bietet. Zudem sollte auch die nahtlose Mobilität dieser Dienste gewährleistet sein.

Das European Telecommunications Standards Institute (ETSI) hat in [8] zwei mögliche Architekturen für ein integriertes Netz spezifiziert, nämlich *enge* und *lose Kopplung*. Diese wollen wir in den folgenden Abschnitten behandeln.

### 1.4.2 Enge Kopplung

Die Idee hinter dieser Variante ist, dass die 3G-Infrastruktur dort, wo sie sich bewährt hat und dem WLAN überlegen ist (Zugangskontrolle, Mobilität, Abrechnungssystem), übernommen werden soll. Das kann erreicht werden, indem das WLAN an einen SGSN angeschlossen wird, wie in Graphik 1.4 gezeigt.

Das WLAN hat dann die Rolle eines weiteren Zugangsnetzes. Die Vorteile dieser Kopplungsart sind:

- Die gemeinsame Abrechnung für beanspruchte Dienste beider Netze ist kein Problem.
- Die Verfahren des 3G-Netzes zur Authentifizierung und zur Verschlüsselung können weiterhin verwendet werden, neu auch im WLAN.
- Es wird möglich, die 3G-Dienste auch aus dem WLAN zu beanspruchen.

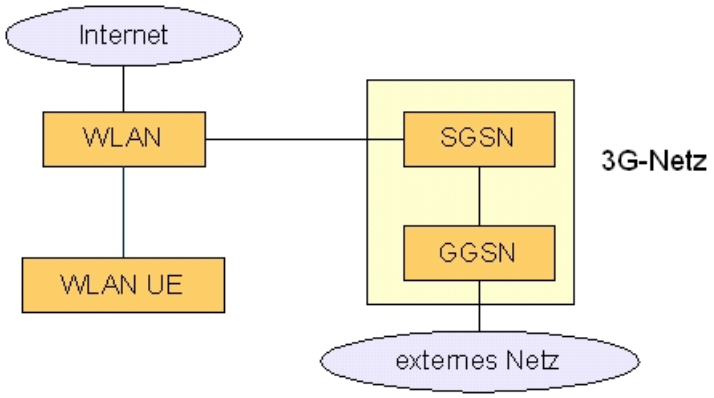


Abbildung 1.4: Enge Kopplung

- Die Fortesetzung der Dienste bei Übergang zwischen WLAN und 3G-Netz ist hier problemlos möglich. Die Benutzer müssen ihre Sitzungen nicht unterbrechen, wenn sie zwischen dem WLAN und dem UMTS-Netz wechseln.
- Durch das Wiederverwenden wesentlicher Teile der Infrastruktur des Mobilfunkbetreibers werden dessen Investitionen besser geschützt und können amortisiert werden.

Damit erfüllt die enge Kopplung die Anforderungen der Szenarien 1 bis 4. Je nach der zugrundeliegenden WLAN-Technologie wird auch Szenario 5 erfüllt (wenn das WLAN eine Dienstqualität unterstützt, die äquivalent ist zu derjenigen der 3G-Netze [12]). Allerdings führt das Prinzip, auf die Infrastruktur des 3G-Netzes zurückzugreifen, auch zu Problemen:

- Bei dieser Art der Kopplung ist es nötig, dass Daten aus dem WLAN das 3G-Kernnetz durchqueren, bevor sie ein externes Netz erreichen. Das bedeutet, dass die enge Kopplung praktisch nur dann realisierbar ist, wenn das WLAN und das 3G-Netz vom selben Betreiber verwaltet werden.
- Die herkömmlichen WLAN-Terminals sind nicht ohne weiteres wieder verwendbar, da sie die UMTS-Protokolle nicht implementieren. Somit macht die enge Kopplung Änderungen an der Struktur der WLANs nötig. Das ist aber kaum zu verwirlichen, da WLANs heute sehr verbreitet sind und viele auch kleinere Firmen WLANs aufbauen und betreuen.

Diese beiden Nachteile sind ziemlich gewichtig, weswegen diese Art der Kopplung in der Regel wenig günstig ist, ausser im Fall, wo das WLAN auch vom Betreiber des 3G-Netzes verwaltet wird.

### 1.4.3 Lose Kopplung

Der andere Ansatz verfolgt das Ziel, möglichst wenige Änderungen an den bestehenden WLANs nötig zu machen. Das WLAN hat hier die Rolle eines externen Datennetzes. Diese

Art der Kopplung wurde in der Technischen Spezifikation des 3GPP im wesentlichen übernommen (siehe [1]). In den folgenden Abbildungen stellen wir diese Art der losen Kopplung vor. In Abbildung 1.5 ist der einfachste Fall dargestellt, wo sich der Benutzer in seinem 3G-Heimnetz aufhält.

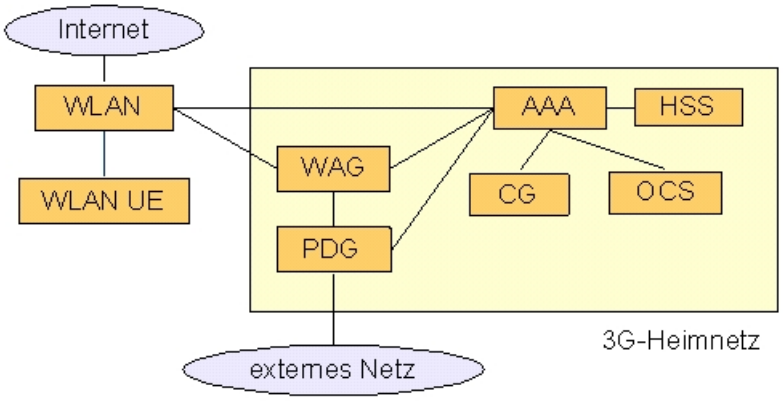


Abbildung 1.5: Lose Kopplung (Benutzer im Heimnetz)

Wenn sich der Benutzer in einem fremden 3G-Netz aufhält (Roaming), so ist das Heimnetz für die Zugangskontrolle verantwortlich. In Abbildung 1.6 sehen wir den Fall, wo die 3G-Dienste vom Heimnetz zur Verfügung gestellt werden.

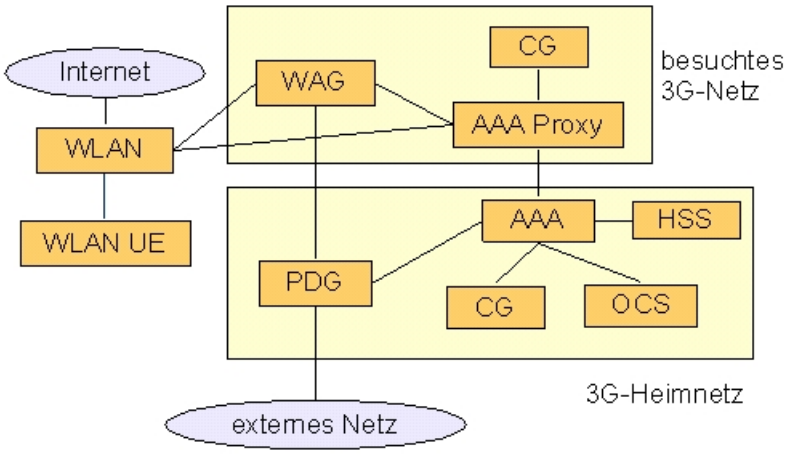


Abbildung 1.6: Lose Kopplung (Benutzer im fremden Netz, erste Variante)

Im Roaming-Fall können die 3G-Dienste auch vom besuchten Netz bereitgestellt werden. Die zugehörige Architektur ist in Abbildung 1.7 dargestellt.

Die Benutzer- oder so genannte *User Equipment (UE)* ist ein Client-Gerät, das eine SIM-Karten Einheit beinhaltet. Es muss kein Simkarten Lesegerät sein, es genügt nur die Unterstützung der UICC/USIM. Das letztere verschlüsselt die Daten, bevor sie übermittelt werden. Zusätzlich hat es eine MS (Mobile Station), das die Funktion der ersten zwei Schichten darstellt und somit die Kommunikation über Funk ermöglicht. Wir werden in den folgenden Abschnitten auch auf die übrigen Komponenten der losen Kopplung eingehen.

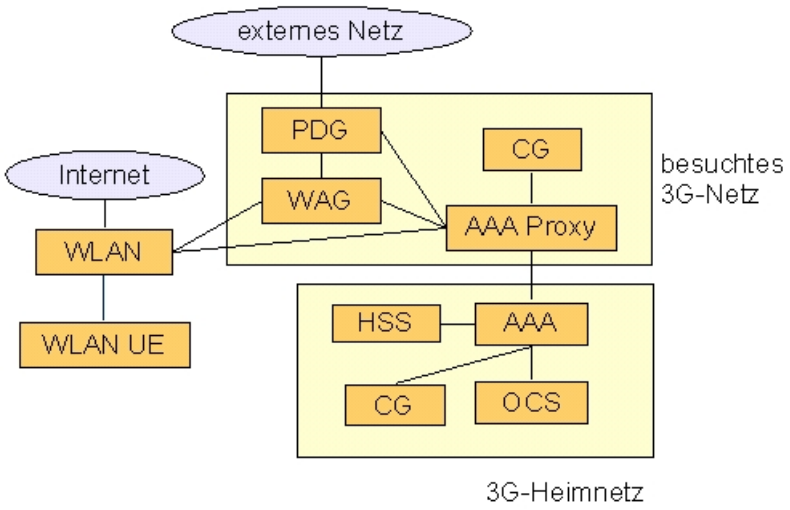


Abbildung 1.7: Lose Kopplung (Benutzer im fremden Netz, zweite Variante)

Wie bereits erwähnt stellt dieses System dem Benutzer zwei neue Möglichkeiten zur Verfügung:

- Zugang, Authentifikation und Autorisation für das WLAN (wird über das 3G-System geregelt)
- Zugang zu externen IP-Netzen (Firmennetz des Betreibers, Intranets oder das Internet) über das 3G-Netz.

Wir werden in den folgende Abschnitten auch genauer sehen, wie diese Funktionen realisiert werden sollen.

Bei der losen Kopplung müssen nur wenige Daten aus dem WLAN das 3G-Kernnetz durchqueren, und die beiden Netze sind weitgehend unabhängig voneinander. Im einzelnen hat sie folgende Vorteile:

- Das WLAN und das 3G-Netz müssen nicht vom selben Betreiber verwaltet werden. Dadurch können auch WLANs, die verschieden aufgebaut sind und verschiedenen Betreibern gehören, mit dem 3G-Netz gekoppelt werden.
- Die heutigen WLAN-Zugangsgeräte brauchen nicht angepasst zu werden. Lose Kopplung basiert auf IETF-Protokollen, die in den WLANs bereits implementiert sind.
- Die lose Kopplung ist mit relativ wenig finanziellem und technischem Aufwand realisierbar.

Andererseits ist es bei der losen Kopplung nicht mehr ohne weiteres möglich, auf die bestehende 3G-Infrastruktur zurückzugreifen. Das führt zu zwei Nachteilen:

- Die gemeinsame Authentifikation, Autorisierung und Abrechnung ist nicht ohne weiteres möglich. Dafür muss der 3G-Betreiber neue Komponenten einsetzen, etwa den AAA-Server. Näheres dazu folgt in Abschnitt 1.5.2.

- Ebenfalls problematisch wird die nahtlose Fortsetzung von Diensten über die Grenzen der Netze hinweg. Es gibt mehrere Ansätze, dieses Problem zu lösen, etwa durch Implementierung des zusätzlichen Protokolls MIP (Mobile IP). Allerdings gibt es dabei Bedenken wegen der Geschwindigkeit — das System könnte zu langsam werden bei Diensten, die eine hohe Übertragungsrate erfordern.

Trotz dieser Nachteile geht der Trend, wie bereits erwähnt, eher in die Richtung der losen Kopplung.

## 1.5 Ausgewählte Aspekte der losen Kopplung

### 1.5.1 Wahl des Netzwerkes

Nehmen wir an, dass sich ein Anwender mit seinem WLAN-Zugangsgerät im Bereich eines WLAN aufhält und sich mit seinem 3G-Heimnetz verbinden will. Weiter gehen wir davon aus, dass der Betreiber des WLANs Roaming-Abkommen mit drei UMTS-Betreibern habe, von denen zwei ein Roaming-Abkommen mit dem Betreiber des 3G-Heimnetzes unseres Anwenders haben sollen, wie in Abbildung 1.8 dargestellt (nach [1]). Das 3G-Heimnetz ist dasjenige Mobilnetz, bei dessen Betreiber der Anwender sein Abonnement hat.

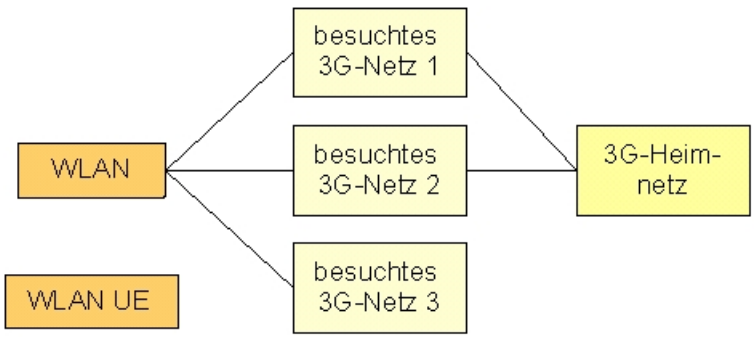


Abbildung 1.8: Netzwahl 1: Welches besuchte Netz wählen?

In der Situation von Abbildung 1.8 kann der Benutzer sein Heimnetz auf zwei Wegen erreichen: entweder über das erste oder das zweite der besuchten 3G-Netze. Das Problem verschärft sich noch, wenn sich unser Anwender nicht nur im Bereich eines einzigen, sondern von mehreren WLANs aufhält, siehe Abbildung 1.9.

Der Betreiber von WLAN 2 hat ein direktes Abkommen mit dem Heimnetz-Betreiber. Es gibt nun in Abbildung 1.9 drei Wege, das Heimnetz zu errreichen: über WLAN 2 direkt zum Heimnetz, oder über WLAN 3 und dann über das besuchte Netz 1 oder 2.

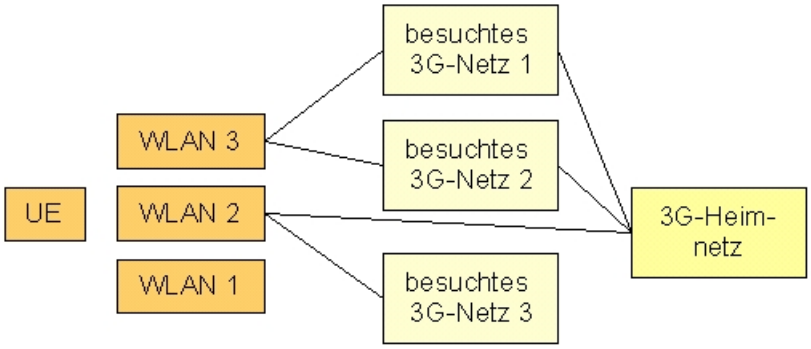


Abbildung 1.9: Netzwahl 2: Welches WLAN und welches besuchte Netz wählen?

*Auswahl des WLAN*: Beim manuellen Modus (Manual Mode) läuft die Auswahl folgendermassen ab: Das Zugangsgerät (UE) erstellt zunächst eine Liste aller WLANs, in deren Bereich es sich befindet (mit Hilfe der SSIDs). Für jedes der erreichbaren WLANs werden dann alle Mobilfunknetze ermittelt, mit denen der WLAN-Betreiber ein Roaming-Abkommen besitzt. Diese Liste der 3G-Netze wird dem Benutzer angezeigt, und er kann manuell einen 3G-Betreiber daraus wählen. Das UE stellt darauf die Verbindung zum WLAN her, das das ausgewählte 3G-Netz unterstützt.

Die Auswahl kann auch automatisch erfolgen (Automatic Mode). Im Zugangsgerät sind dann zwei Listen gespeichert: eine enthält die vom Heimnetz-Betreiber bevorzugten SSIDs und die andere die vom Benutzer bevorzugten SSIDs. Die erste enthält etwa die SSIDs derjeniger WLAN-Betreiber, mit denen der Betreiber des 3G-Heimentzes ein Roaming-Abkommen besitzt. Das Zugangsgerät scannt dann diese Listen durch. Details zu diesem Algorithmus findet man in [1].

*Auswahl des 3G-Netzes*: Nachdem sich das Zugangsgerät mit dem ausgewählten WLAN verbunden hat, sendet es einen Authentication Request in Form eines NAI (Network Access Identifier) an das WLAN. Eine NAI hat das Format Benutzername@Bereich, wo der Bereich wie der Name einer Internet-Domäne aufgebaut ist. Das WLAN versucht dann, diese Anforderung an das in der NAI angegebene 3G-Netz weiterzuleiten. Wenn das nicht gelingt (da der Betreiber des im Bereich angegebenen 3G-Netzes kein Abkommen mit dem WLAN-Betreibe hat), sendet das WLAN die Roaming-Partner und ihre NAI-Bereiche an das Zugangsgerät. Dieses verarbeitet die erhaltenen Daten nach seinen internen Roaming-Richtlinien und präsentiert dem Benutzer eine Liste, aus der er ein zu besuchendes 3G-Netz auswählt. Aufgrund dieser Auswahl wird eine neue NAI erstellt, mit der sich das Zugangsgerät erneut zu authentifizieren versucht usw.

### 1.5.2 Sicherheit, Authentifikation und Autorisierung

Bei jeder neuen Architektur, handelt sich dabei um existierende Komponenten, die zusammengefügt werden wie in unserem Fall, oder um neue Technologien, spielt die Sicherheit eine grosse Rolle bei der Entwickelung, Implementierung und bei der Inbetriebnahme. Wenn man von Sicherheit redet, so trifft man immer wieder folgende Begriffe: Datenschutz, Authentifikation, Vertrauenswürdigkeit und Integrität. Alle Begriffe adressieren

immer die gleichen Anforderungen, nur gibt es einige Unterschiede bezüglich der drahtlosen Kommunikation.

Unter Datenschutz in der drahtlosen Kommunikation versteht man die Sicherstellung des Schutzes der Position eines Clients im System. Es ist nämlich leicht, diese beiden Entitäten zu assoziieren.

Authentifikation bezieht sich auf die Identifikation der Systemzustände und Prozesse mit einem Client. Der Client authentifiziert sich beim System und verlangt eine gewisse Mobilität. Was bedeutet, dass er sich frei aus einem System in das andere bewegen will, ohne sich immer wieder authentifizieren zu müssen.

Vertrauenswürdigkeit ist die Sicherstellung des Schutzes von Benutzer- und Systemdaten, sowohl stationär als auch bei der Übertragung. In 3GPP-WLAN wird dies mit symmetrischer Verschlüsselung erreicht.

Integrität bedeutet Schutz der Daten vor Manipulation. Es wird anhand von Message authentication codes (MAC) erreicht. MACs gehören zu den symmetrischen Verschlüsselungsverfahren. Asymmetrische Verfahren wären zu teuer bezüglich den Ressourcen.

Bevor wir die Sicherheit der Integrierten Architektur anschauen, werden wir im Folgenden die Sicherheit der WLAN und der UMTS Technologie betrachten.

WLANs gelten als unsicher. Die Gründe hierfür sind vielfältig. Erstens wäre da der fehlende Standard für den Zugriff und den Aufbau eines WLAN Systems. Zweitens sind die Funkwellen, wie oben genannt, sehr leicht abhörbar, da sie eben nicht isoliert durch Kabel reisen, sondern in der Luft. Mit WEP hat man versucht, der leichten Abhörbarkeit entgegenzuwirken. WEP steht für Wired Equivalent Privacy und ist ein Verschlüsselungsverfahren, welches mit dem RC4 Algorithmus arbeitet. Der Schlüssel für den Algorithmus besteht aus zwei Teilen:

1. WEP-Schlüssel fixer Länge.
2. Initialisierungsvektor

Leider ist der Algorithmus geknackt worden. Dies liegt daran, dass es sich bei RC4 um einen linearen Algorithmus handelt. Solche gelten als unsicher. Die Schwachstelle wurde soweit ausgenutzt, dass mit der Kenntnis einer Klartextnachricht beliebige mit WEP verschlüsselte Nachrichten erstellt werden können [16].

Eine weitere Möglichkeit, um die Sicherheit in WLAN zu erhöhen, stellt die VPN (Virtual Privat Networks) dar. Nur entspricht diese nicht den Anforderungen der Provider von Mobilen Kommunikation, da VPN ressourcenaufwändig sind.

Bei UMTS sieht es schon etwas anders aus bezüglich der Sicherheit. Die Netze der zweiten und dritten Generation gelten als besonders sicher und verlässlich. Die Basis der UMTS Sicherheit ist die Authentication und Key Agreement (AKA) Prozedur. Die Vorteile von AKA sind netzwerkunabhängig und können über verschiedene Mechanismen laufen. AKA

benutzt die Simkarte in UMTS Geräten, die die Daten verschlüsseln während der Ausführung von AKA Prozedur.

Die bestehenden Sicherheitsmechanismen sind im Link Layer implementiert. Konkret bedeutet das, dass jegliche Versuche, um Sicherheitsstandards der Drahtlosenkommunikation zu verbessern, über dieser Schicht stattfinden müssen.

Zusätzlich besteht ein grundlegender Unterschied bezüglich der Sicherheit zwischen den lose und eng gekoppelten Architekturen, nämlich die Anforderungen an die WLAN Architektur. Betrachtet man eine enge Kopplung, so ist eine Veränderung des grossteils der 802.11 Spezifikation erforderlich. Hingegen verlangt die lose Kopplung nur die Implementierung der Authentifikation Funktion. Wir erinnern uns daran, dass alle 802 Standards über einen einheitlichen Link Layer verfügen müssen, dass somit alle darüber liegenden Protokolle auf die gleichen Dienste und Schnittstellen zugreifen können.

Die Bestandteile der Authentifikations-Funktion sind das Extensible Authentication Protokoll (EAP) und das AAA-Protokoll. Diese agieren unter anderem als Transportmechanismen über den Link Layer.

Aus den oben beschriebenen Sicherheitsaspekten und -mechanismen lassen sich Anforderungen an die 3GPP-WLAN Architektur ableiten.

1. UMTS Sicherheitsarchitektur bildet die Basis für die Sicherheit in 3GPP-WLAN Architektur.
2. Die WLAN Systemarchitektur wird um eine Komponente erweitert, welche den AAA und EAP unterstützt.
3. Zusätzlich sollten die Mobilen Clients Zugriff auf die Simkarten haben. Genauer sollte ein Zugriff auf die Verschlüsselungsverfahren der Simkarte möglich sein.

Da die Sicherheitsarchitektur von UMTS Technologie beibehalten wird, fallen keine Kosten für die Schlüsseldistribution an.

Im Folgenden sollen die Protokolle der UMTS Technologie genauer angeschaut werden. Jedoch müssen vorher Entitäten in einem 3GPP-WLAN System und deren Beziehungen genau definiert werden.

Man unterscheidet zwischen drei fundamentalen Komponenten im System:

- Bereits früher erwähnt wurde die Benutzer-Ausrüstung (UE).
- *Home Environment (HE)* ist das Herz des Systems und enthält den Home Subscriber Server (HSS), welches die Benutzerdaten speichert, und den AAA Server.
- *Serving Network (SN)* ist WLAN Netzwerk des Systems und besitzt einen AAA Poxy Server, den Network Access Server (NAS), der eine Anzahl von Access Points im Netzwerk kontrolliert und die Access Points.

Wie bereits festgehalten, die wichtigste Voraussetzung für die Sicherheitsprotokolle ist das AAA Protokoll, das auf RADIUS oder DIAMETER Protokollen basiert.

**Authentication and Key Agreement Protocol**

In folgender Grafik ist das Vorgehen bei einem AKA Protokoll dargestellt.

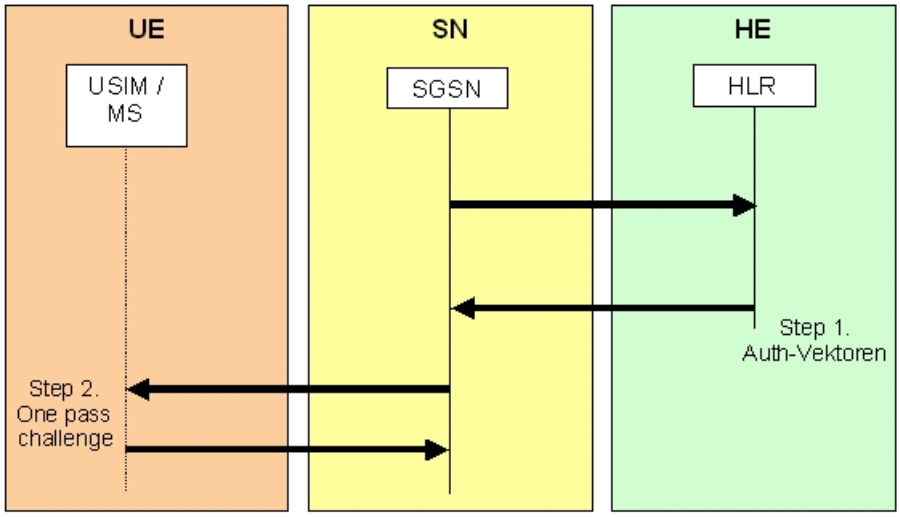


Abbildung 1.10: Übersicht über die AKA-Prozedur

Der AKA Protokoll läuft in zwei Phasen ab. In der ersten Phase tauschen die HE und SN Einheiten die Daten, die notwendig sind für eine Authentifikation, in Form eines Authentifikation-Vektors aus. Dies geschieht automatisch und hängt nicht direkt von der Anmeldung des UE ab. In der zweiten Phase verlangt die UE Einheit den Zugang zu WLAN. Sie wird automatisch erkannt und es werden Daten von SN an UE gesendet. Dabei handelt es sich um einen Challenge, der aus einer Zufallszahl und einem Authentication Token erstellt wird. Der UE errechnet die Antwort RES und sendet sie zurück. SN verifiziert die RES und entscheidet ob die Prozedur erfolgreich beendet worden ist oder nicht.

Es können einige Fehler passieren während dem Vorgang, wie z.B. dass es zu einem MAC Fehler kommt oder die Sequenznummer der Nachrichten nicht stimmt.

EAP geht einen Schritt weiter als das AKA-Protokoll. Es agiert als ein Transportmechanismus und bietet gleichzeitig mehrer Authentifizierung Mechanismen. Der grosse Vorteil von EAP ist, dass es auf dem Link Layer läuft und keine Erweiterungen der WLAN Architektur erfordert. Zusätzlich ist es möglich, auf andere Authentifikation Verfahren zurückzugreifen, auch wenn sie nicht von EAP unterstützt werden. In so einem Fall assistiert das EA-Protokoll bei der Authentifikation mit dem HSS (Home Subscriber Server).

Nun schauen wir uns den EAP-AKA Protokoll im Detail an (siehe Abbildung 1.11) und wir werden auf die wichtigsten Schritte eingehen:

1. Es wird die Identität des UE auf Anfrage von SN / WLAN ausgetauscht.
2. SN sendet die erhaltene Nachricht an den AAA Server im HE.
3. Dieser leitet es weiter an die HSS. HSS überprüft die Identität, bzw. die Existenz des Benutzers und senden die Sicherheitsdaten an den AAA Server zurück

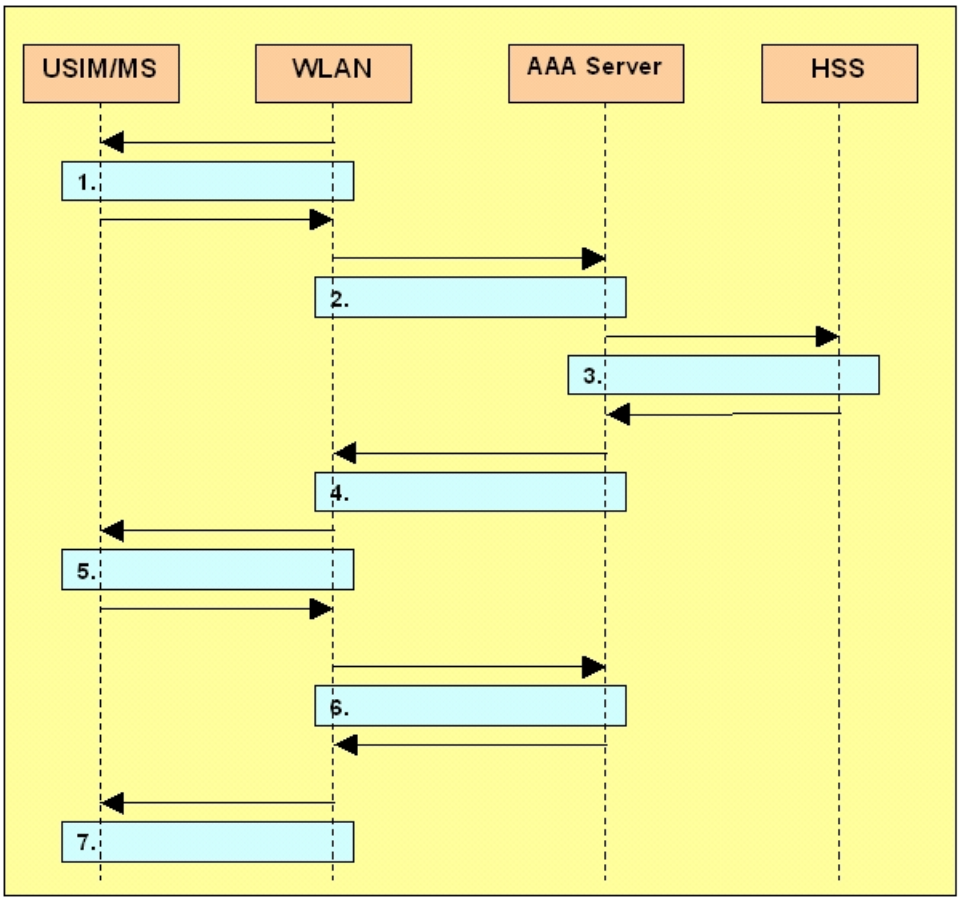


Abbildung 1.11: Übersicht über die EAP-AKA-Prozedur

4. Der AAA Server sendet den Challenge, erstellt anhand von den Sicherheitsdaten, an den SN / WLAN.

5. SN / WLAN sendet diese zurück an den UE, dieser berechnet den Response und sendet es zurück an den SN.

6. SN / WLAN sendet diesen zurück an den AAA Server, der den Response überprüft und entsprechend der Korrektheit den Schlüssel erstellt und zurückschickt oder abweist.

7. Der UE wird nur mit SUCCESS oder ERROR benachrichtigt über die Gültigkeit.

Integrität und der Vertrauenswürdigkeit Mechanismen in dem Link Layer erklärt die 3GPP gemäss dem Paper [3] nicht als Forschungsthema und empfiehlt die Mechanismen vom 802.11i Spezifikation. Diese Spezifikation definiert den höheren Sicherheitsstandard durch den Einsatz von einem neuem Verschlüsselungsverfahren, so genannten WPA.

Noch zu erwähnen sind die Integrität und Vertrauenswürdigkeit Mechanismen der UMTS Technologie. Hierbei handelt es sich um zwei Verschlüsselungsalgorithmen, den UEA (UMTS Encryption Algorithm) und den UIA (UMTS Integrity Algorithm). Die Verschlüsselung geschieht in der MAC oder RLC Schicht, die der Link Layer Control Schicht entspricht. In dieser Arbeit soll nicht genauer auf die beiden eingegangen werden.

### 1.5.3 Dienstqualität

Da die Anfragen an die Provider der UMTS Technologie immer grösser werden, die populäre WLAN zu integrieren, so steigt auch die Notwendigkeit eines Quality of Service Teilsystems. Der Grund hierfür ist die Eigenschaft von UMTS Netzwerken, QoS für IP Kommunikation anzubieten (Beispiel Voice over IP). Um dies weiterhin anbieten zu können, insbesondere für die End-zu-end Paketschaltung sprich IP / WLAN Kommunikation, so muss QoS Management in die integrierte Architektur für UMTS / WLAN übernommen werden.

Am Anfang werden wir uns mit einigen Definitionen auseinandersetzten um später auf die Details einzugehen.

Quality of Service (QoS):

[14] „Quality of Service (QoS) refers to the capability of a network to provide better service to selected network traffic over various technologies.... The primary goal of QoS is to provide priority including dedicated bandwidth, controlled jitter and latency (required by some real-time and interactive traffic), and improved loss characteristics. Also important is making sure that providing priority for one or more flows does not make other flows fail.“

[15] „Bei der Auslegung eines Datennetzes, das auch für Voice-over-IP genutzt werden soll, muss die Laufzeit, die Daten im Netzwerk haben, möglichst niedrig gehalten werden. Um das zu erreichen, werden Datenpakete, die Sprache enthalten, bei der Weiterleitung in Routern und Switches der Vorzug gegeben. Hier werden so genannte Class-of-Services(CoS)- und Quality-of-Services(Qos)-Mechanismen eingesetzt.“

Aus diesen zwei Definitionen ist ersichtlich, um was es sich handelt bei QoS. UMTS definiert vier verschiedene QoS Klassen:

- *background* für Datentransfer mit möglichst geringer Fehlerrate aber unkritischen Anforderungen an Bandbreite, Delay und Jitter.
- *interactive* für die Nutzung interaktiver Dienste. Ähnliche Anforderungen wie bei Background jedoch höhere Anforderungen ans Delay, um Wartezeiten bei der Nutzung der Dienste zu vermeiden.
- *streaming* für Verteildienste. Eine Mindestbandbreite ist erforderlich; Jitter ist in gewissem Rahmen erlaubt, da empfängerseitig Jitterbuffer verwendet werden. Bitfehler sind eher unkritisch.
- *conversational* für direkte Kommunikation (Telephonie, Videotelephonie). Ähnliche Anforderungen wie Streaming, aber deutlich weniger Jitter und Delay sind zulässig.

Die Klassen sind jedoch an die QoS Policen gebunden. Dieser werden von den jeweiligen Providern definiert und beinhalten die Regeln und Bestimmungen, anhand welchen der Provider die entsprechenden QoS Klassen in seinem System garantiert.

In der Architektur des QoS Management Systems werden diese Regeln und Bestimmungen der Policen in Policy decision point (PDP) oder Policy decision functionality (PDF) gespeichert und verwaltet. Die AP der WLAN Technologie werden mit einer weiteren Komponente erweitert, die so genannte Policy enforcement point (PEP). PEP steht im folgenden Zusammenhang mit PDP: PEP stellt eine Anforderung an PDP. PDP übersetzt die Policenregeln in Mechanismen anhand des Netzzustands und der Fähigkeit von PEP. Die Mechanismen werden an PEP gesendet und von dieser ausgeführt.

Das Verlangen an Mobilität bezieht sich nicht nur auf einen Provider sondern auf mehrere. In solchen Fällen muss das System mit mehreren QoS Policen klarkommen. Die UMTS Technologie setzt hierfür einen interdomain policy agent ein, der die Policen eines Providers mit anderen Providern updatet. Vor dem Update werden die Policen jedoch in die eigenen Regeln übersetzt und dann repliziert.

In lose gekoppelten Integrationen sind die WLANs über die GPRS Gateways mit UMTS verbunden. Im Bezug auf QoS Dienst weist diese ebenfalls einen grossen Vorteil auf, nämlich den, dass es ohne grossen technischen Aufwand möglich ist, die WLAN Netzwerke samt ihren Policen in andere Netzwerke zu integrieren. Verständnishalber ist zu erwähnen, dass die WLAN Technologie auch über einen QoS Dienst verfügt. Dieser wird von den AP und WLAN Router gemanaged.

Genauso wie in UMTS Netzen können auch in integrierten Architekturen mehrere Provider und zusätzlich mehrere WLANs eingebunden sein. Um möglichst viele Fälle abzudecken, werden im Folgenden drei allgemeine Fälle definiert und analysiert.

**Fall 1. Beide Technologien werden vom gleichen Provider administriert**

In diesem Fall betreibt der Provider die gesamte integrierte Architektur. Das System verfügt über ein hierarchisch aufgebautes Policen System. Dies bedeutet, dass die QoS Policen von WLANs den Policen des UMTS untergeordnet sind. Der Master PDF (MPDF) verbindet die WLAN PDF (WPDF) und den UMTS PDF und übersetzt gleichzeitig die Policen einzelner WLANs in die Gesamtarchitektur Policen. Diese werden dann in der zentralen Policen Datenbank gespeichert.

Im einzelnen werden die folgenden Schritte durchlaufen:

1. WPDF erhält die QoS Parameter für eine Sitzung von einer AP / WR.
2. Danach holt sich die WPDF die Policen Regeln von der zentralen Policen Datenbank.
3. Nachdem die Regeln erhalten worden sind, werden die QoS Parameter der Sitzung überprüft, ob sie den Regeln entsprechen. Sind die Parameter richtig, so wird die AP benachrichtigt und Police umgesetzt, sind die Parameter falsch, so wird AP über die Nichtgültigkeit benachrichtigt.
4. Kommt es zu einem Konflikt bezüglich den Policen z.B. wegen Authentifizierung Regeln, so stellt die WPDF eine Anfrage an MPDF um den Konflikt zu lösen. Dies geschieht mit Hilfe eines Protokolls genannt Common Open Policy Service (COPS).

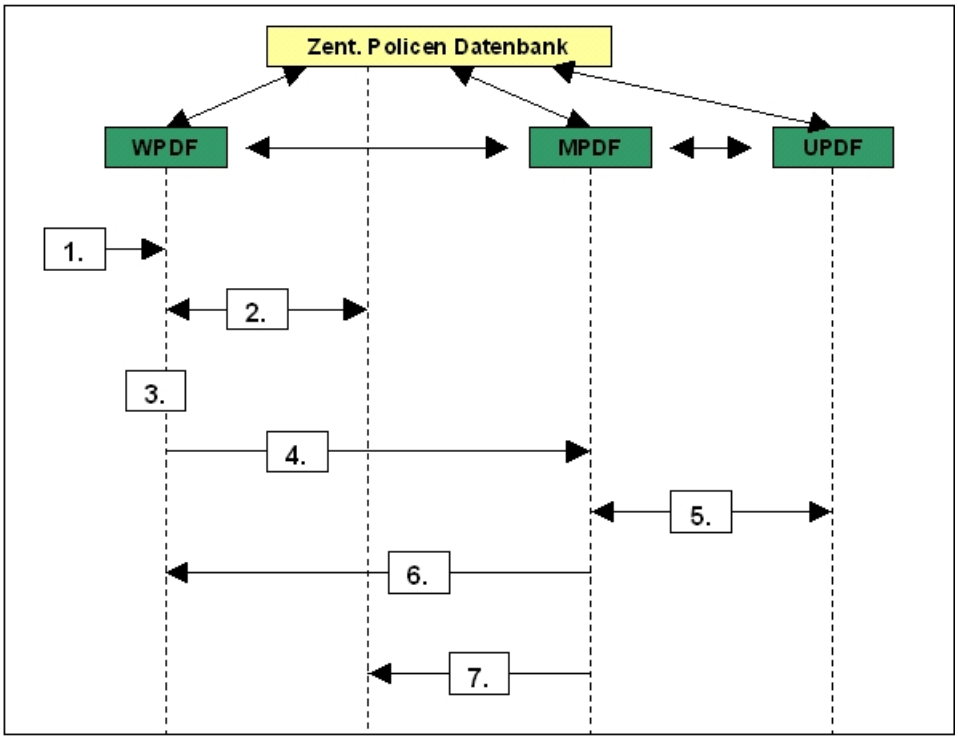


Abbildung 1.12: Policen-Management beim gleichen Provider

5. MPDF errechnet neue Police anhand der Daten aus UMTS PDF, die für die Session bestimmt ist.

6. Ist die Prozedur beendet, so sendet MPDF die Police an WPDF ...

7. ... und macht einen Update in der zentralen Policen Datenbank.

**Fall 2. Ein WLAN Netz wird von verschiedenen Providern geteilt**

Aufgrund des Roaming Problems in WLAN Technologie und der Kostenfrage ist es üblich dass mehrere verschiedene Provider eine WLAN Architektur aufbauen und sich diese teilen. So können sich Provider mit verschiedenen Grundpolicen das gleiche WLAN teilen. Ein Beispiel: ein Provider garantiert die besten Bandbreiten und Services während der Arbeitszeit während ein anderer dies ausserhalb der Arbeitszeit tut. Somit können auch Probleme bezüglich den QoS Policen entstehen, da der WPDF für jede Verbindung die entsprechenden MPDF anfragen muss. Logischer Weise können hier Konflikte entstehen bezüglich den Policen. Man hat zwei Wege entwickelt um dieses Problem zu lösen, erstens die WPDF überlässt es den verschiedenen MPDF die Policen auszuhandeln oder sie implementiert eine überschreibende Police im solchen Fällen. Da die WLANs über die WPDF bereits verfügen, ist es kostengünstiger die zweite Lösung zu wählen.

Wie man in der Grafik sieht geschieht die Policen Aushandlung in einigen Schritten mehr.

1. AP / WR senden die QoS Parameter an die WPDF.

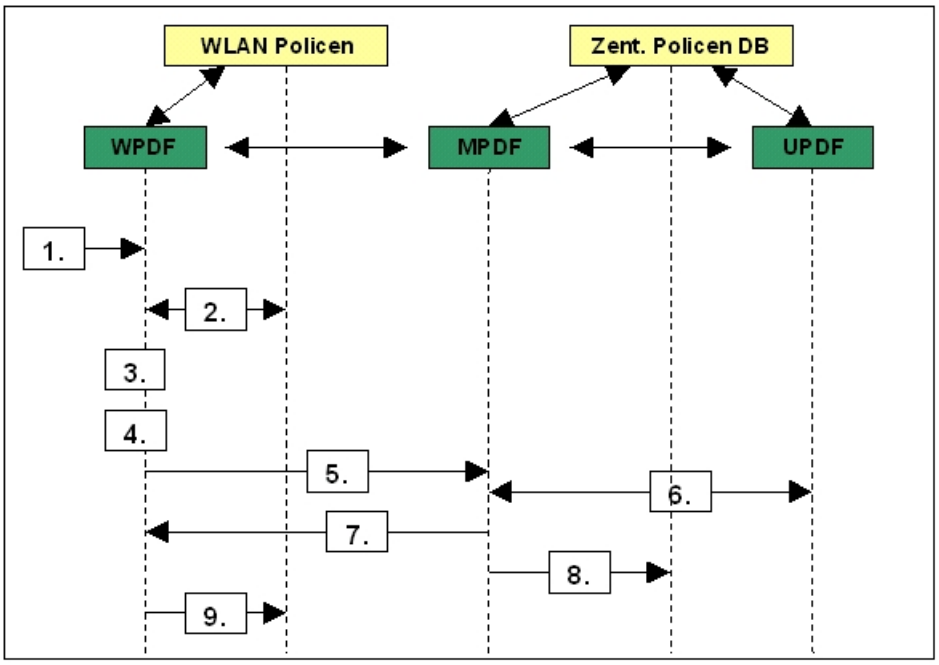


Abbildung 1.13: Gleiches WLAN, mehrere Provider

2. WPDF holt sich die Policen von WLAN Policen Datenbank.

3. WPDF kontrolliert die Parameter ob sie erlaubt sind. Falls ja, wird die Sitzung der Policenregeln nach gemanaged. Falls nicht, dann wird die Anfrage abgelehnt.

4. Im Falle eines Konfliktes sendet die WPDF die Parameter an die MPDF um das Problem zu lösen. Falls es einen zusätzlichen Konflikt bezüglich den Policen der einzelnen Provider gibt, so löst WPDF selber diesen Konflikt mit der überschreibenden Police.

5. Die Parameter werden in den COPS Protokoll verpackt und an die MPDF gesendet.

6. MPDF erstellt eine neue Dienstgüte anhand der Parameter und den Policen, die zusätzlich noch von den UMTS PDF überprüft werden.

7. Die neue Dienstgüte wird an WPDF zurück geschickt.

8. MPDF updatet die zentrale Police Datenbank.

9. WPDF wendet die Dienstgüte an und updatet eigene Policen DB.

### Fall 3. WLAN gehört einem Provider, der Kunde ist beim UMTS Provider

Das WLAN gehört einem Provider, der ein Kunde eines UMTS Netzes ist. Dies erlaubt dem UMTS Provider ein grosses Gebiet mit WLAN Zugang abzudecken. Der grundlegendste Unterschied zu den vorherigen Fällen, ist der, dass WPDF das alleinige Recht zu einem WLAN Policen Update hat. Da aber mehrere UMTS Provider in die Kommunikation involviert sein könne, sind die Konflikte im Bezug auf die Policen vorprogrammiert.

Nun da sich nur die WPDF selbst updaten dürfen, entsteht ein Problem bei der Aushandlung der Policen. Um dieses Problem zu lösen, wird eine neuer Mechanismus eingeführt, Service Level Specifications (SLS). Sie werden von den beteiligten Parteien ausgehandelt und definieren die zu implementierende Dienstgüte. Die Aushandlung kann statisch und dynamisch sein. Statische können direkt aus den einzelnen Policen übersetzt und eingesetzt werden, die dynamischen werden jedoch von allen beteiligten Netzen ausgehandelt. Dabei erlaubt die SLS die Aushandlung über die IPA (Interdomain policy agents). Ist die Aushandlung abgeschlossen, so übersetzen die IPAs die Police in ihren eigenen Netzwerken und setzen sie um.

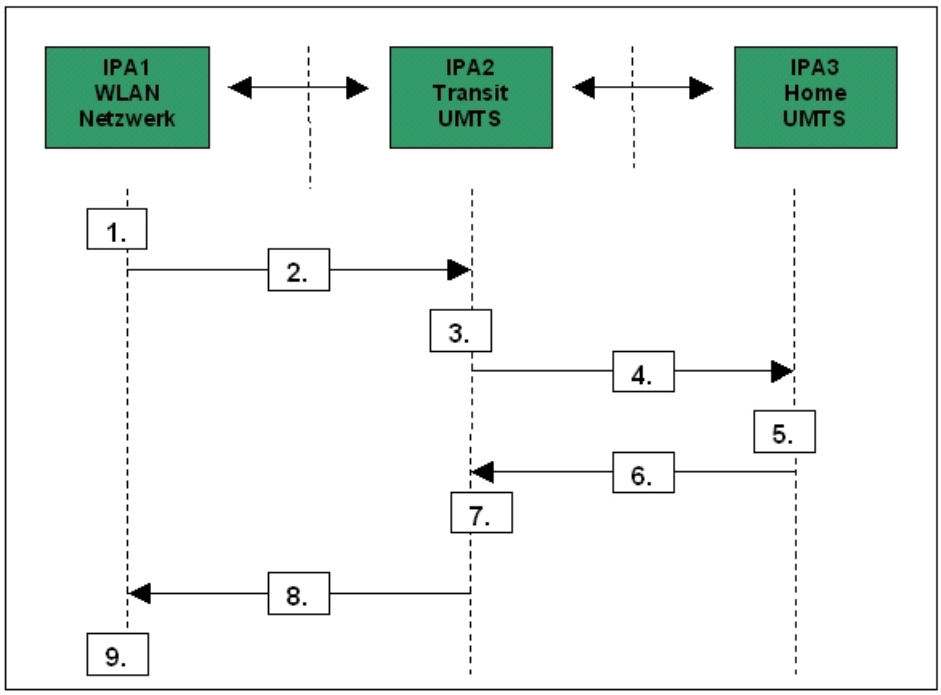


Abbildung 1.14: SLS Policen-Aushandlung

Es wird wie folgt vorgegangen:

1. WPDF entdeckt einen Fall, dass den aktuellen Policen nicht entspricht. Die Parameter werden in SLS Protokoll geschrieben

2. IPA1 verpackt den SLS Protokoll in eine COPS Anfrage und sendet sie zum Transit UMTS IPA2.

3. IPA2 entpackt die Daten und konsolidiert eigenen MPDF ob die Police realisierbar ist. Falls ja, so wird eine positive Antwort an den WPDF zurückgeschickt. Der Grund: Transit UMTS garantiert, dass die Police auch vom Home UMTS übersetzbar ist.

4. Falls MPDF des Transit UMTS die Police nicht übersetzen kann, so wird sie an den IPA3 im Home UMTS weitergeleitet.

5. Schritt 3 wird wiederholt....

In Schritten 6 bis 9 wird die COPS Antwort (positiv / negativ) zurückgesendet und die eigenen Policen werden updatet durch die neue Dienstgüte, jedoch nur falls die Antwort positiv ist. Falls die Antwort negativ ist, findet kein Update statt und die Anfrage kann mit veränderten Parametern von IPA1 erneut gestartet werden.

Wie wir gesehen haben, decken die Szenarien eine hierarchische Architektur (Fall 1), eine Peer to Peer (Fall 3) und eine Mischform ab. Diese drei Architekturbeispiele sollten abstrakt genug sein um die meisten Einzelfälle abzudecken.

### 1.5.4 Datenfluss und Abrechnung

Zunächst betrachten wir den Zugang zu externen Datennetzen über das 3G-System. Dieser kann über das 3G-Heimnetz des Benutzers und — je nach Abonnenment und Roaming-Abkommen — auch über ein besuchtes 3G-Netz erfolgen. Im 3G-Netz bildet ein *WAG* (WLAN Access Gateway) die Schnittstelle für paketbasierte 3G-Dienste. Diese Station müssen (bei einem paketbasierten Dienst) alle Daten passieren, die von einem externen Netz in das WLAN oder umgekehrt gelangen wollen.

Nachdem sich der Benutzer über das WLAN authentifiziert und autorisiert hat, so übermittelt das WLAN-Zugangsgerät mit Hilfe des sogenannten APN (Access Point Name) den Dienst, den der Benutzer beanspruchen will. Das 3G-Netz ermittelt daraus den zuständigen *PDG* (Packet Data Gateway). Dieser ist ähnlich wie ein GGSN und bildet die Schnittstelle zwischen dem 3G-Netz und einem äusseren Datennetz. Zwischen dem WLAN-UE und dem PDG wird ein Tunnel eingerichtet, d. h. eine sichere Punkt zu Punkt-Verbindung. Durch diesen Tunnel können dann die Daten übermittelt werden. Dasselbe Zugangsgerät kann auch mehrere externe Verbindungen gleichzeitig unterhalten, dann werden dementsprechend mehrere Tunnels eingerichtet.

Abgesehen von seiner Funktion als Gateway erfüllt der WAG noch weitere Zwecke: Er sorgt dafür, dass die Datenpakete aus dem WLAN durch den entsprechenden PDG geroutet werden. Umgekehrt stellt er sicher, dass nur Pakete von diesem zuständigen PDG in das WLAN gelangen. Zudem sammelt er (jedenfalls im Roaming-Fall) auch für jeden durch ihn führenden Tunnel Informationen für die Abrechnung, etwa nach der Anzahl der übertragenen Bytes oder nach der abgelaufenen Zeit.

Die Abrechnung geschieht in beiden Fällen (direkter Internet-Zugang und Dienst via ein externes Netz) über das 3G-System. Bei der Abrechnung gibt es grundsätzlich zwei Varianten: prepaid oder postpaid. Beim *prepaid*-System muss der Benutzer zuerst eine Anzahl Kreditpunkte kaufen, und er erhält ein Konto, das vom *OCS* (Online Charging System) verwaltet wird. Wenn unser Benutzer dann einen Service beansprucht, werden die verbrauchten Kreditpunkte von diesem Konto online abgezogen.

Beim *postpaid*-System hat der Benutzer einen Vertrag mit dem Betreiber, der die Abrechnungsdaten sammelt und dann dem Benutzer periodisch (etwa monatlich) eine Rechnung zustellt. Die Daten für die Abrechnung werden im *CG* (Charging Gateway) gespeichert und für das Ausstellen der Rechnung in bearbeiteter Form an das Rechnungssystem weitergegeben.

*Abrechnung für Internet-Zugang über das WLAN:* Bevor der Benutzer Zugang zum WLAN erhält, wird im prepaid-Fall der AAA-Server die nötigen Kreditpunkte im OCS reservieren. Falls diese auf dem Konto des Benutzers vorhanden sind, schickt der OCS eine Bestätigung an den AAA-Server, und der Benutzer kann autorisiert werden. Die Abrechnungsdaten werden vom WLAN an den AAA-Server weitergeleitet, der diese Daten in einem bestimmten Format (als WLAN CDRs, WLAN Access Detail Records) an den CG weiterleitet. Diese Informationen werden dann verwendet, um die Abrechnung zwischen den Betreibern des WLAN, des Heimnetzes und gegebenenfalls der besuchten Netze vorzunehmen. Ausserdem werden diese Daten im postpaid-Fall auch zum Erstellen der Rechnung verwendet.

*Abrechnung von IP-basierten 3G-Diensten:* Wie erwähnt durchqueren alle IP-Pakete einen PDG, und daher werden die Abrechnungsdaten dort gesammelt. Bei prepaid-Benutzern fordert der PDG — ähnlich wie der AAA-Server vorher — vom OCS die erforderlichen Kreditpunkte an, und er muss auch sicherstellen, dass die gebrauchten Dienste den Wert des Kredits nicht überschreitet. Der PDG gibt die Abrechnungsdaten ebenfalls an den CG weiter, wo sie analog wie beim Internetzugang über WLAN für die Abrechnung zwischen den Betreibern und für die Ausstellung der Rechnung an postpaid-Benutzer verwendet werden.

## 1.6 Zusammenfassung und Fazit

UMTS und WLAN sind zwei Technologien, die sich für die Übertragung von Sprache und Daten eignen. Ihre Vor- und Nachteile verhalten sich komplementär, woraus sich die Frage ergibt, ob nicht beide in einem integrierten Netz kombiniert werden könnten, damit so ihre Vorteile zum Tragen kommen. Es wurden zwei Varianten behandelt, wie ein solches integriertes Netz aussehen könnte: enge und lose Kopplung. Bei der engen Kopplung hat das WLAN die Rolle eines Zugangsnetzes, und die Vorteile des 3G-Netzes sollen möglichst unverändert übernommen werden. Das Problem dabei ist, dass diese Art der Kopplung praktisch nur dann in Frage kommt, wenn das WLAN und das 3G-Netz demselben Betreiber gehören. Bei der losen Kopplung sollen beide Netze möglichst eigenständig bleiben, was allerdings einige zusätzliche Komponenten nötig macht wie einen AAA-Server oder einen PDG. Damit sollte es möglich sein, die Verfahren des 3G-Netzes für die Authentifikation, Autorisierung und Abrechnung im wesentlichen auch im integrierten Netz zu übernehmen. Ausserdem haben wir diskutiert, wie die Dienstqualität (QoS) in den verschiedenen Fällen sichergestellt werden kann. Die lose Kopplung scheint sich durchzusetzen, da sie einfacher zu realisieren ist.

Grundsätzlich geht der Trend in Richtung eines Systems, das den verlässlichen und sicheren Zugang zu vielfältigen IP-basierten Diensten ermöglicht, und zwar über verschiedene Zugangstechnologien, deren Komplexität dem Benutzer verborgen bleibt [4]. Im Moment ist allerdings noch nicht in Sicht, dass ein solches System in der Schweiz bald in Betrieb genommen wird — jedenfalls hat Swisscom auf eine entsprechende Anfrage per E-Mail nicht reagiert, und auf ihrer Homepage war dazu nichts zu finden.

# Literaturverzeichnis


[1] 3GPP: 3GPP System to Wireless Local Area Network (WLAN) Interworking, TS 23.234 v6.2.0, September 2004.

[2] 3GPP: Feasibility Study on 3GPP System to WLAN Interworking, TR 22.934 v1.2.0, Mai 2002.

[3] 3GPP: 3G Security; WLAN Network interworking security, 2004

[4] K. Ahmavaara, H. Haverinen, R. Pichna: Interworking Architecture Between 3GPP and WLAN Systems, IEEE Communications Magazine, November 2003.

[5] J. Ala-Laurila, J. Mikkonen, J. Rinnemaa: Wireless LAN Access Network Architecture for Mobile Operators, IEEE Communications Magazine, November 2001.

[6] D. I. Axiotis, T. Al-Gizawi, K. Peppas, E. N. Protonotarios: Services in Interworking 3G and WLAN Environments, IEEE Wireless Communications, October 2004.

[7] Bundesamt für Kommunikation (BAKOM): „Faktenblatt“ UMTS, Version 2.2, 16. November 2004.

[8] ETSI: Requirements and Architectures for Interworking between HIPERLAN/3 and 3rd Generation Cellular Systems, TR 101 957, August 2001.

[9] P. Lescuyer: UMTS, dpunkt.verlag, 2002.

[10] J. Roth: Mobile Computing, Grundlagen, Techniken, Konzepte; dpunkt Verlag; 2002

[11] A. K. Salkintzis: Interworking Techniques and Architectures for WLAN/3G Integration Towards 4G Mobile Data Networks, Motorola.

[12] A. K. Salkintzis, C. Fors, R. Pazhyannur: WLAN-GPRS Integration for Next-Generation Mobile Data Networks, IEEE Wireless Communications, October 2002.

[13] S. M. Syed: An Analysis of 3G-WLAN Integration, ENTS 609.

[14] http://www.cisco.com/univercd/cc/td/doc/cisintwk/ito_doc/qos.htm#wp1024961, März 2005.

[15] http://www.elektronik-kompendium.de/sites/net/0905131.htm, März 2005.

[16] http://www.wikipedia.de, März 2005.

# Kapitel 2

# DVB-basierte Rundfunksysteme

*Alex Müller, Bernhard Wasser*

*Dieser Seminarbericht gibt einen Überblick zu den existierenden digitalen Rundfunkstandards Digital Audio Broadcasting (DAB), Digital Radio Mondiale (DRM) und Digital Video Broadcast (DVB). DVB wird weiter klassifiziert in DVB-T, DVB-H, DVB-S und DVB-C. Die potentiellen Anwendungsgebiete der Standards werden aufgezeigt sowie deren tatsächlicher Stand der Implementierung und Verbreitung. Im technischen Teil wird auf das Komprimierungsverfahren MPEG-2 als Grundlage für die hohen Datenraten beim digitalen Rundfunk und die Funktionsweise der DVB Standards eingegangen. Die Versendung von IP Paketen über Rundfunknetzwerke mit Hilfe des generischen IPDC Frameworks wird erklärt. MBMS ist ein erstes Derivat einer Anwendung von IPDC. Das IPDC Framework ist die Grundlage für neue IP basierte Services via Broadcast-Netzwerke. Satellite Digital Multimedia Broadcast (S-DMB) ist eine Entwicklung, die Satelliten in die mobile Datenübertragung einbindet. Abschliessend werden die vorgestellten Standards einander zum Vergleich gegenübergestellt und die Konvergenz im Mobilbereich beleuchtet.*

## Inhaltsverzeichnis

## 2.1 Rundfunktechnologien: Ein Überblick

Im Zuge der Digitalisierung und den Ansprüchen der Gesellschaft nach Mobilität, stand bisher die Bastion des analogen Rundfunks noch ohne zu wanken. Von einigen digitalen PayTV Angeboten abgesehen, sah es bis zur Jahrhundertwende nicht so aus, als würde sich in Europa digitales Fernsehen in Kürze durchsetzen. Im Radiobereich tat sich auch nichts. Obwohl die Standards eigentlich schon existierten, gab es kaum Angebote auf dem Markt, weder Empfänger noch Programme. Doch in den letzten Jahren zeigten sich verschiedene Standards, die das Potential besitzen, sich durchzusetzen. In diesem Seminarbericht werden unter anderem drei solche digitale Standards beleuchtet: Digital Audio Broadcast (DAB), Digital Radio Mondiale (DRM) und Digital Video Broadcast (DVB).

Diese drei Standards haben einiges gemeinsam: Alle dienen dem Zweck digitalen Rundfunk zu verbreiten [2] [9] [4]. Alle werden in Europa von verschiedenen non-profit Konsortien stark gefördert und ihnen ist der bisherige Durchbruch noch nicht wirklich gelungen. Obwohl zu jedem Standard schon Angebote bestehen, wankt das analoge Imperium noch nicht. Die drei hier vorgestellten Standards könnten zwar Daten auch auf mobile Geräte übertragen, dennoch zielt jede Technologie auf eine andere Nische ab, weil sie sich im potentiellem Angebot und Übertragungsrate stark unterscheiden.

### 2.1.1 DAB

Das Digital Audio Broadcasting - kurz DAB - ist ein Standard zur Übertragung von Rundfunk in digitaler Form. Entwickelt wurde dieser Standard für die Europäische Union von EUREKA, einem Europäischen Konsortium aus Broadcastern, der Industrie und Forschungsinstituten. EUREKA wurde im Jahre 1985 gegründet. Ziel von EUREKA ist das marktorientierte Forschen und Entwickeln im Technologiesektor. Der DAB Standard - auch Eureka 147 genannt - wurde im Zeitraum von 1987 bis 2000 entwickelt, mit dem Ziel den analogen Rundfunk bis ins Jahre 2010 gänzlich abzulösen [1].

Im Gegensatz zum bisherigen analogen Radio verspricht DAB Klangqualität auf dem Niveau einer Audio-CD. Gerade im Bereich des mobilen Empfangs verfügt DAB über grosses Potential, da auch bei hohen Fortbewegungsgeschwindigkeiten ein stabiles und interferenzfreies Signal übermittelt werden kann. Die Übertragung erfolgt terrestrisch, sprich ohne Satellit; für den Empfang benötigt man eine entsprechende, ungerichtete Antenne. DAB basiert auf der MPEG-2 Codierung (vgl. Kapitel 1.3.2): Die Datenströme von ursprünglich über 2000 KBit/s werden auf 32 bis 256 KBit/s reduziert, dann übereinander geschichtet und moduliert; man spricht hier von einem Multiplex. Doch nicht nur reine Audioströme können zu einem Ensemble zusammengefasst werden, auch Datendienste können zugleich mit übertragen werden [2].

#### Verwendung

Die verwendete Modulationstechnik ermöglicht Ausstrahlungen, die deutlich robuster als mit der bisherigen analogen Technik sind. Zudem ist es möglich, weite Flächen mit nur

einer Frequenz abzudecken, was dazu führt, dass man nicht ständig die Frequenz am Radio anpassen muss, wenn man über weite Strecken unterwegs im Auto oder im Zug ist. Möglichkeiten für übertragene Datenströme sind beispielsweise Radiotext für zusätzliche Informationen zu den gerade gespielten Songs oder aber Zusatzmeldungen wie Stau- und Unfallwarnungen. So ist es den Anbietern auch möglich, die Kategorie ihrer übertragenen Musik in Stichworten anzugeben, was eine Kanalsuche mittels Schlagworten wie "Jazz" oder "Rock" seitens des Empfängers erlaubt - ähnlich dem Schema bei der Suche nach passenden Web-Radiostationen. So existieren auch schon viele Geräteanbieter, die ihre neuen Autoradios mit einem DAB-Empfänger ausrüsten, da die Sender langsam damit beginnen, auch ihr DAB-Angebot zu verbessern. Gerade das mangelnde Angebot an Empfängern war in den vergangenen Jahren eines der Hauptprobleme bei der Verbreitung von DAB.

#### Verbreitung

Das Ausbreitungsgebiet von DAB beschränkt sich momentan noch auf den Europäischen Kontinent. Das European Telecommunications Standards Institute (ETSI) hat DAB als europäischen Standard festgeschrieben; die USA und Japan haben jeweils ihre eigenen Standards entwickelt. Der geplante Ablösungstermin im Jahre 2010 rückt aber mehr und mehr in die Ferne, da obwohl in Deutschland schon eine 80% Abdeckung gewährleistet wird, eher eine Abkehr von DAB seitens der Programmanbieter zu verzeichnen ist [2]. Jedoch in Bayern, wo im Jahre 2000 auch der erste deutschlandweite DAB-Sender startete, und in der Schweiz wird der DAB Programmausbau immer noch sukzessive vorangetrieben. In der Schweiz hat die SRG (Schweizerische Rundfunk Gesellschaft) eine führende Rolle eingenommen. So bietet sie mehrere Sender an, darunter auch DRS1 bis DRS3, Virus, Radio Swiss Jazz, Radio Swiss Pop sowie verschiedene italienische und französische Radioprogramme [40]. Auch für die Jahre 2005 und 2006 ist ein weiterer Ausbau geplant, da man DAB bei der SRG grosses zukünftiges Potential einräumt.

Europaweit konnten bisher ungefähr 12 Millionen Empfänger abgesetzt werden. Doch noch bestehen nicht in allen europäischen Ländern auch DAB Angebote. Bisher existieren solche Übertragungen primär in Skandinavien, England, dem deutschen Sprachraum und in Portugal. Länder wir Frankreich, Italien aber auch weite Gebiete im Osten sind erst noch in der Planungs- und Testphase (siehe Abbildung 2.1) [3]. Wann analoges Radio endgültig der Vergangenheit angehören wird, ist noch nicht absehbar.

### 2.1.2 DRM

Digital Radio Mondiale, kurz DRM, steht für einen nicht proprietären Standard für weltweiten Rundfunk auf den Lang-, Mittel- und Kurzwellenfrequenzen. Die Codierung basiert auf der MPEG-4 Technologie. Das DRM-Projekt entstand im September 1996 bei einem informellen Treffen von verschiedenen grossen internationalen Rundfunkanstalten in Paris. Vertreten waren unter anderem Radio France Internationale, Deutsche Welle, Voice of America und noch ein zwei weitere Stationen und Gesellschaften, die zu den Grossen dieser Branche gezählt werden können; die offizielle Gründung erfolgte dann am 5. März 1998 in China. Unterdessen umfasst dieses non-profit Konsortium nicht nur Broadcaster,

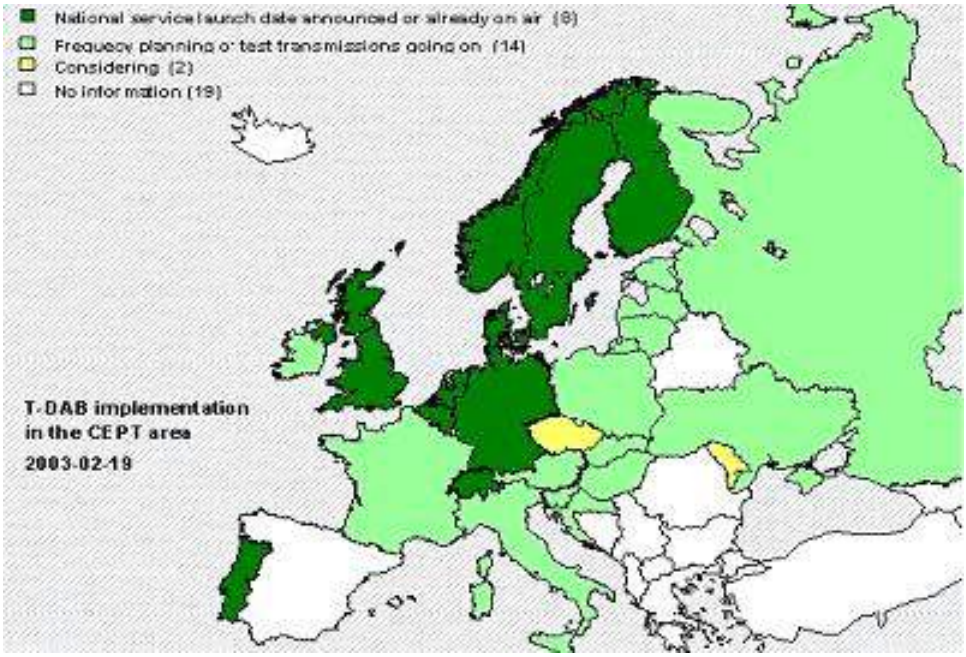


Abbildung 2.1: Verbreitung von DAB in Europa. [3]

auch Forschungsinstitute und Netzwerkbetreiber zählen sich zu den mittlerweile 80 Mitgliedern.

Das Ziel des DRM Projektes ist es, mit DRM einen Standard zu schaffen, der den momentan recht unpopulären analogen Rundfunk auf besagter Lang-, Mittel- und Kurzwelle ablösen kann. Da es sich bei DRM genau wie bei DAB um einen nicht-proprietären Standard handelt, sind die gesamten Spezifikationen auch auf der Website des ETSI kostenlos verfügbar [4].

### Verwendung

Die Qualität des heutigen analogen Rundfunks ist relativ gesagt bescheiden. Dies liegt an der verwendeten Modulationstechnik, die dazu führt, dass nur ein Monokanal übertragen werden kann. In diese Bresche will nun DRM einspringen. Die Qualität von DRM ist einiges besser, kommt aber nicht an UKW - oder gar DAB-Qualität heran. Gesendet wird auf der selben Wellenlänge, die Radiostationen müssten sich also keine neue Frequenz zulegen, wenn die Umstellung beginnt. Ähnlich wie bei der DAB-Technologie kann auch mit DRM nicht nur Radio, sondern auch Text und Daten übertragen werden, sei es im Sinne von Zusatzinformationen zum Programm oder auch Metadaten zu den jeweiligen Musikstücken. Die Programme können mobil - sowohl via Autoradio wie auch via portable Radios und PDAs - empfangen werden, die korrekte Empfangsantenne vorausgesetzt [5] [6].

Momentan existieren zwar schon einige Empfänger, die meisten sind jedoch relativ teuer und bisweilen auch recht unhandlich. Die zweite Generation an portablen Empfängern

- entwickelt von der Firma Mayah - ist seit letztem Jahr erhältlich und ist schon einiges kompakter. Zudem existieren verschiedene Angebote für den Empfang via PC, meistens mittels einer USB Steckkarte. Für den Broadcaster kommt der Betrieb dieser neuen Technologie günstiger: "die Energiekosten liegen bei gleicher Reichweite etwa 40 Prozent niedriger als beim klassischen AM-Sender" [7].

#### Verbreitung

Das DRM Angebot ist bisher relativ klein, bislang senden 70 Anbieter ihre Programme aus. Manche nur an bestimmten Wochentagen, die meisten Stationen - etwa 50 - senden aber bereits täglich. Informationen über einen Schweizer Sender sind auf der offiziellen Sendeliste nicht auffindbar. Immerhin senden verschiedene Anbieter deutschsprachige Programme, wie beispielsweise RTL Radio. Hörer die mehr auf exotisches Programm aus sind, können sich mit Radio Sweden, Radio China, Radio Kuweit und Radio Vatikan die Zeit vertreiben. Die meisten Sender sind europäischer Herkunft, aber es existieren auch Anbieter in den USA, Afrika, Russland und dem mittleren Osten. Sogar China hat sich zu DRM bekannt, und ist laut Zhang Haitao, dem Vize-Minister der Staatsverwaltung für Radio-Rundfunk, "auf dem Wege, DRM für Mittelwelle/AM und Kurzwelle zu nutzen" [8].

### 2.1.3 Vergleich DAB - DRM

Da es sich mit DAB und DRM um zwei verschiedene Rundfunk-Projekte von zwei verschiedenen Konsortien handelt, stehen sie nicht in direkter Konkurrenz. Sie unterscheiden sich sowohl beim Kundensegment, wie auch beim Frequenz-Spektrum und kommen sich deshalb nicht direkt in die Quere. Verschiedene Gerätehersteller bieten gar Radioempfänger an, die den Empfang beider Technologien ermöglichen. Potential und Qualität sind weitere Unterscheidungspunkte im Wettbewerb der zwei Standards: Ein einzelner DAB-Sender versorgt ein lokal begrenztes geografisches Gebiet mit Musik in CD-Qualität und einem vielseitigen Programmangebot auf täglicher Basis; ideal für Leute die einfach nur Musik hören wollen oder ihrem lokalen Sender treu sind. DAB-Nutzer wollen auch unabhängig von der Tageszeit mit möglichst kleinem Aufwand und viel Komfort mobil auf möglichst viele Programme in HiFi-Qualität zugreifen können.

Lang-, Mittel- und Kurzwellen Programme dagegen zielen mehr auf die Versorgung weiter Flächen und ganzer Staaten oder auf die Ausstrahlung in weit entfernte Gebiete ab. Oft wird auch nur zu speziellen Tageszeiten oder gar nur an bestimmten Wochentagen gesendet; das Programm ist meistens auch entsprechend exotisch. Die Programmanbieter setzen voraus, dass der Hörer bereit ist, einen gewissen Aufwand zu treiben, um in den Genuss der Sendung zu kommen - beispielsweise bei der Sprache oder bei ungewöhnlichen Sendezeiten. Dies alles gilt jetzt auch für die DRM Anbieter, da sie ja auf das selbe Hörersegment ausgerichtet sind. DRM wird wohl nie der Rundfunk für die Masse werden, da die Qualität doch hörbar schlechter ist als bei DAB.

Ob sich beide Standards dauerhaft verbreiten und durchsetzen können, ist noch nicht absehbar. Rein von der Nachfrage hat DAB sicher die Nase vorn, nur schon wegen dem

Tabelle 2.1: Vorteile und Nachteile von DAB und DRM.

| DAB | DRM |
|---|---|
| + hohe Klangqualität | + Frequenzen bleiben |
| + frequenzökonomisch | + geringer Stromverbrauch |
| + tägliches Programm | - mittelmässige Qualität |
| + kleine Antennen | - geringe Auswahl |
| + ideal für mobile Geräte | - spezielle Sendezeiten |
| - nur Europa | - kleines Hörersegment |

besseren Angebot. Sollten die Gerätehersteller in Zukunft sinnvolle Geräte auf den Markt bringen, die sowohl DAB wie auch DRM gleichzeitig empfangen und die jeweiligen Metadaten auch wirklich nutzen können, wird einer friedlichen Koexistenz kaum etwas im Wege stehen. Im Gegenteil, die Hörer können gar vom verbreiterten Angebot profitieren, und jeweils bequem nach Präferenzen den geeigneten Kanal für sich finden [4].

### 2.1.4 DVB

Digital Video Broadcasting, kurz DVB, ist im Bereich Radio einer der Hauptkonkurrenten von DAB. Doch wie es der Name schon andeutet, ist mit DVB nicht nur Radioübertragung möglich, sondern jegliche Datenübertragung; der Hauptanwendungsbereich liegt jedoch in der Übertragung von Filmen und Videosequenzen. DVB bezeichnet das standardisierte Verfahren zur Übertragung von digitalen Inhalten wie Fernsehen, Radio, Mehrkanalton oder interaktiven Diensten wie Teletext und elektronische Programmübersicht, die zusätzlich zum Fernsehprogramm übertragen werden können.

Durch Datenkompression - momentan noch MPEG-2, in Zukunft auch MPEG-4 oder eine ähnliche Technologie - können im Vergleich zur analogen Fernsehübertragung mehr Programme pro Frequenzband übertragen werden [9]. Die Qualität des übertragenen Inhalts lässt sich beliebig anpassen: Je stärker die Daten komprimiert werden, desto mehr Programme können miteinander Übertragen werden. Im Gegenzug stinkt jedoch die verfügbare Qualität und der Rechenaufwand steigt. Das Signal lässt sich relativ sicher und kostengünstig verschlüsseln, was für PayTV unbedingt nötig ist, da die Anbieter ja davon leben, dass nicht jeder ihr Programm anzapfen kann, ohne dafür zu bezahlen. Es existieren mehrere Unterarten des DVB-Standards; die vier wichtigsten sind DVB-S, DVB-C, DVB-T und DVB-H [9] [10].

#### DVB-S

DVB-S ist die Abkürzung für Digital Video Broadcast - Satellite. DVB-S ist eine spezielle Variante von DVB, die via Satellit übertragen wird. Die grundlegende Technik ist dieselbe, aber diese Variante enthält Optimierungen für die satelliten-spezifische Eigenschaften bei der Übertragung von digitalen Daten, wie beispielsweise ein Fehlerkorrekturverfahren. Die Ausstrahlung von DVB via Satellit ist heute noch die meistgenutzte DVB Variante, da die

grosse Bandbreite die Übertragung der meisten Fernseh- und Rundfunkprogramme sowie Zusatzdienste zulässt.

Im Vergleich zu anderen DVB Varianten wird auch keine Zusatzinfrastruktur benötigt, da die Satelliten bereits im Weltraum schweben [34]. Dies macht es dann auch möglich, dass man als Konsument von DVB-S keine zusätzlichen Gebühren zahlen muss - der Satellitenbetrieb wird von den Sendeanstalten bezahlt. Da die Miete eines Satelliten-Transponders recht kostenintensiv ist, profitieren auch die Sendeanstalten von DVB-S, da damit ja mehr Programme pro Transponder ausgestrahlt werden können im Vergleich zur analogen Technologie. Die zweite Generation von DVB-S - das sogenannte DVB-S2 - ist dank verbesserter Codierung bis zu 30% effizienter. Die Spezifikationen liegen dem ETSI vor, und es wird damit gerechnet, dass dieser Standard den alten bald ablösen wird [9].

### DVB-C

Das C steht für Cable - die Übertragung erfolgt also über den Kabelanschluss. Nur wer auch am Kabelnetzwerk angeschlossen ist, kommt in den Genuss von dieser DVB-Variante. Obwohl die Kabelabdeckung in der Schweiz im internationalen Verglich hoch ist, verfügen längst nicht alle Haushalte über Kabel. In abgelegenen Gebieten existieren bisweilen gar keine Anbieter, in andern Regionen nur solche mit übertriebenen Preisen. Grundsätzlich ist das mögliche Angebot das selbe wie bei anderen DVB-Angeboten.

Im Vergleich zur DVB-S Variante ist die Bandbreite jedoch eingeschränkt. Das digitale Kabelangebot steckt im deutschsprachigen Gebiet - im Vergleich zu den USA - noch in den Kinderschuhen. Neben den öffentlich-rechtlichen Sendern, die ihr Programm neu auch digital anbieten, und dem PayTV-Anbieter Premiere gibt es nur ein überschaubares Angebot im Kabel, meist in Form kostenpflichtiger ausländischer Programmpaketen [9].

### DVB-T

DVB-T ist die Abkürzung für Digital Video Broadcasting - Terrestrial, die Übertragung erfolgt terrestrisch über Antennen. Dieser Standard ist primär in Europa verbreitet; Japan und die USA entwickelten jeweils ihre eigenen Standards, die auch auf MPEG-2 basieren. Mit DVB-T kann bei gleicher Reichweite mit weniger Leistung gesendet werden als bei analogem Rundfunk, was die Infrastrukturanbieter natürlich freut. Mit DVB-T könnte man in der gesamten Schweiz mit derselben Frequenz senden, ohne dadurch Interferenzen in Kauf nehmen zu müssen. Momentan sieht die Frequenz-Landschaft noch sehr heterogen aus, da aus technischen Gründen nicht zwei Gebiete mit derselben Frequenz aneinander anliegen dürfen. Für den Empfang wird auch bei dieser Variante ein Receiver benötigt, eine sogenannte Set-Top-Box. Diese Empfangsgeräte sind momentan noch etwas teuerer als die Pendants von DVB-S und DVB-C, dafür ist keine weitere Installation wie eine Satellitenschüssel notwendig. Für den terrestrischen Empfang fallen im Gegensatz zum Empfang via Kabel nur geringe laufende Kosten an, die man auf den Zuschauer abwälzen könnte [9].

### DVB-H

DVB-H ist die Abkürzung für Digital Video Broadcasting - Handhelds. In der Literatur werden bisweilen auch die Namen DVB-M und DVB-X benutzt, die früher für die Bezeichnung dieses Standards gängig waren. Mit DVB-H können Fernsehübertragungen auf mobile Endgeräte wie Mobiltelefone oder Laptops übertragen werden. Genau wie DVB-T wird auch bei DVB-H terrestrisch über Antenne gesendet; die Technik baut deshalb auch auf derer von DVB-T auf.

DVB-H soll die Technologie werden, die zum ersten mal TV-Übertragungen in guter Qualität auf mobile Endgeräte ermöglichen soll, speziell auf Mobiltelefone. Ziel ist es - möglichst Batterie schonend - TV-Übertragungen und weitere DVB-Dienste zu übermitteln, die auch bei hoher Geschwindigkeit des Endbenutzers noch ohne grosse Störungen empfangen werden können. Zum effektiven Einsatz von DVB-H wird das IP Datacast System benötigt, das sich derzeit noch im Spezifizierungsprozess befindet, und später in dieser Abhandlung detaillierter durchleuchtet wird [9].

### Verbreitung

Hinter dem europäischen DVB-Projekt stehen über 260 Mitglieder, die sich zu einem Konsortium zusammengeschlossen haben, um das digitale Fernsehen mit dem DVB-Standard weiterzubringen [10]. Darunter befinden sich auch kanadische und australische Firmen, sowie japanische und amerikanische Unternehmen, die via ihre Tochtergesellschaften am DVB-Projekt beteiligt sind. Die Mitglieder kommen aus verschiedensten Branchen, hauptsächlich sind Programmanbieter, Gerätehersteller, Netzbetreiber und europäische Behörden beteiligt. Wie auch bei den Radiostandards ist auch hier das ETSI wieder vertreten, was dazu führt, dass auch die DVB-Spezifikationen kostenlos verfügbar sind. Auf der Karte kann man erkennen, dass DVB ähnlich verbreitet ist wie auch schon DAB.

Am weitesten mit der Einführung ist England, wo bereits verschiedene grössere Angebote bestehen. Aber auch Skandinavien, Deutschland und die Schweiz planen momentan wie man das Angebot ausbauen, und auf das ganze Land ausbreiten kann (vgl. Abb. 2.2). Das europäische Parlament hat den Nutzen von DVB erkannt und hat im Oktober 2002 eine Resolution verabschiedet, die sich für eine rasche Verbreitung von DVB in Zukunft einsetzt. Nicht nur Europa gehört jedoch ins gewünschte Zielgebiet, auch Afrika und der mittlere Osten sollen in das Einzugsgebiet von DVB gehören. Ob Japan und die USA ihre eigenen Standards aufgeben, und ein gemeinsam entwickelter, kompatibler Standard einmal herrschen wird, ist momentan noch nicht absehbar [3].

## 2.2 Anwendungsbereich

Da DVB sich mehr und mehr auf dem Vormarsch befindet, ist es sinnvoll, mögliche Verwendungszwecke aufzuzeigen. Grundsätzlich kann man zwischen zwei verschiedenen Kategorien von DVB-Diensten unterscheiden: IP-based und TV bezogene Dienste. Die IP

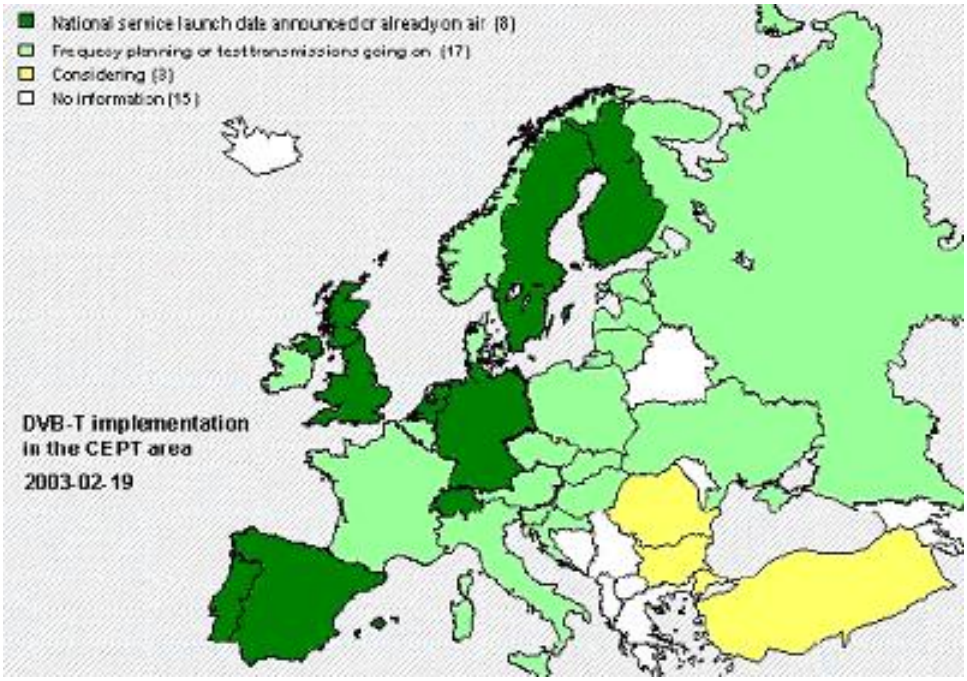


Abbildung 2.2: Verbreitung von DVB-T in Europa. [3]

basierten Dienste zielen hauptsächlich auch die Datenübermittlung auf mobile Endgeräte ab, Fernsehübertragungen sind da auch mit eingerechnet.

### 2.2.1 IP-basierte Datendienste

Die Vielfalt an möglichen Diensten ist gross, einige bestehen auch schon mittels anderen Technologien. Obwohl heute die mobilen Dienste via Mobiltelefon oder PDA noch nicht ausschöpfend genutzt werden, wird die Nachfrage in Zukunft wohl auch in Europa zunehmen, wenn Dienste erscheinen, die auch wirklich auf Akzeptanz bei den Nutzern stossen. Um die Masse an verschiedenen IP basierten Diensten zu kategorisieren, werden sie in drei Untergruppen aufgeteilt: Streaming-, Interaktive- und Nicht-Interaktive-Dienste [11].

#### Streaming Services

Streaming und der Download von Audio und Video unterscheidet sich darin, dass Streaming Dateien in der Regel nicht beim Nutzer abgespeichert werden. Der User bestellt sich einen speziellen Dienst, den er entweder gleich oder zu einem späteren Zeitpunkt konsumiert - er bezahlt jedoch nur das einmalige Anschauen der Datei. Will er dieselbe Datei zu einem späteren Zeitpunkt nochmals sehen, muss er sie ein weiters mal downloaden; je nach Dienst natürlich auch noch einmal bezahlen. Aus diesem Grund sind Streaming-Angebote oft relativ günstig, da man relativ wenig von der Datei hat, und auch keine Gefahr der unerlaubten Weitergabe besteht. Videostreams sind oft stark komprimiert und dadurch nur von mittelmässiger Qualität, besonders wenn sie auf ein mobiles Endgerät übermittelt

werden. Die Qualität hängt dabei stark von der möglichen Übertragungsrate und der verfügbaren Batterieleistung ab. Wenn die Bandbreite und die Batteriepower stimmt, kann ziemlich jede Bildqualität übertragen werden. Aber gerade der Batterieverbrauch ist momentan noch ein Problem, das schnell gelöst werden muss; denn von einem Mobiltelefon das nur an der Ladestation hängt, hat der Nutzer wenig [11].

Anbieten könnte man beispielsweise Liveübertragungen von Ereignissen, die sich einer grossen Interessengruppe erfreuen. Denkbar wären Sportereignisse, die Fussballspiele am Samstagnachmittag oder für Radsport-Fans die tägliche Zielankunft an der Tour de France - alles über das eigene Mobiltelefon. Eine andere mögliche Angebotssparte wären Livekonzerte. All die armen Seelen die mal wieder keine Eintrittskarten fürs U2-Konzert ergattern konnten, hätten so immerhin die Möglichkeit von überall her das Konzert live mit zu verfolgen. Auch die Übertragung des regulären TV-Programmes auf die mobilen Geräte würde sich sicher grosser Beliebtheit erfreuen, gerade bei Pendlern oder anderen Leuten die viel auf Achse sind. Der nächste Schritt wäre dann TV-on-demand oder wenigstens near-on-demand - der Anwender sucht sich selber von einer Liste möglicher Übertragungen aus, was er sehen und wann er es sehen möchte [11].

### Interactive Services

Um interaktive Services anbieten zu können benötigt man eine Zweiwegkommunikation. Ein Antwortkanal ermöglicht es dem Nutzer auf den ihm angebotenen Inhalt zu reagieren - je nach Dienst können auch Fremdtechnologien benutzt werden um einen solchen Antwortkanal zu realisieren. Momentan werden SMS (Short Message Service) oder UMTS (Universal Mobile Telecommunications System) als Antwortkanäle in Betracht gezogen. Mobiles Online-Shopping wäre ein realisierbarer Interaktiver Dienst: Der Kunde greift mittels Mobiltelefon auf den Katalog des Anbieters zu, sucht sich mittels SMS oder einem anderen Medium seine gewünschten Waren zusammen in seinen Warenkorb, und bezahlt die dann mit Kreditkarte. Obwohl heute solche Angebote nur begrenzt genutzt werden, sollte man nicht ausschliessen, dass die Nachfrage in diesem Sektor in Zukunft steigt. Ein weiteres Beispiel für einen interaktiven Dienst wäre ein DVB-Chat. Da die übermittelte Datenmenge bei einem Chat relativ gering ist, liesse sich dies gut realisieren. Eine weitere Variante wären Multiplayer-Online-Spiele, bei denen man direkt mit seinem mobilen Gerät gegen andere Spieler im ganzen Land antreten kann. Kleinere Spiele könnten über eine direkte Verbindung zwischen den Nutzern stattfinden. Bei grösseren Anwendungen lässt man das Geschehen auf einem Server laufen, auf den die Mitspieler online zugreifen [11].

Das klassische Beispiel für einen interaktiven Dienst bleibt das bestellen von Kinokarten: Irgendwo auf dem Lande wird beschlossen den Abend mit einem Kinofilm zu verbringen, doch anstatt sich jetzt zum nächsten PC mit Internetanbindung zu begeben, zückt unsere Versuchsperson einfach seinen PDA und greift auf das Angebot des lokalen Kinos zu. Nachdem verschiedene Trailer per Streaming angeschaut und bewertet worden sind, wird der Wunschfilm auserwählt, und die besten noch verfügbaren Sitze aus dem Sitzplan angewählt. Die Bestätigung wird abgeschickt und direkt mit der Kreditkarte beglichen.

### Non-Interactive Services

Ein Antwortkanal ist bei diesen Diensten nicht notwendig. Entsprechend eingeschränkt ist deshalb natürlich das Angebot - die Dienste dienen hauptsächlich zur Information des Nutzers. Gängige Angebot wären Börsenticker für die Geschäftswelt, Sportresultate, Wettvorhersagen und Nachrichten-Ticker. Pure Datenübertragung ist natürlich auch möglich mittels DVB - das entsprechende Gerät muss dazu mit einem DVB Empfänger ausgerüstet sein. Wenn sowohl der PC, das Laptop, der DVD-Recorder, der MP3-Player und die Spielkonsole untereinander vernetzt sind, steigt das Heimkino auf eine neue Stufe auf [11].

### Charakteristika von Diensten

Um einen erfolgreichen Dienst zu lancieren, müssen zuerst einige Charakteristiken für den Service festgelegt werden. Ein wichtiger Kernpunkt ist die Ausbreitung des Dienstes. Soll nur regional Verfügbar sein, oder das ganze Land abdecken? Wer darf auch den Dienst zugreifen? Möglichkeiten wären eine Abonnementstruktur, pay-per-use oder aber man macht den Dienst kostenlos und öffentlich zugänglich, und verdient selber nur an der plazierten oder übermittelten Werbung. Wie genau die Zahlung dann erfolgt, ist ein weiteres Problem, das noch durchdacht werden muss. Wenn der Inhalt nicht jedem zugänglich sein soll, sollte man eine Verschlüsselung in Betracht ziehen, um den Zugang einzuschränken. Ob der Inhalt ständig verfügbar sein soll, und ob er auf dem Endgerät gespeichert werden soll, hängt ganz vom Inhalt ab. Ob und welche Dienste sich in ferner Zukunft etablieren können, wird sich noch zeigen [11]. Abbildung 2.3 ordnet mobile Dienste nach Eignung im Spannungsfeld zwischen Rund- und Mobilfunk ein.

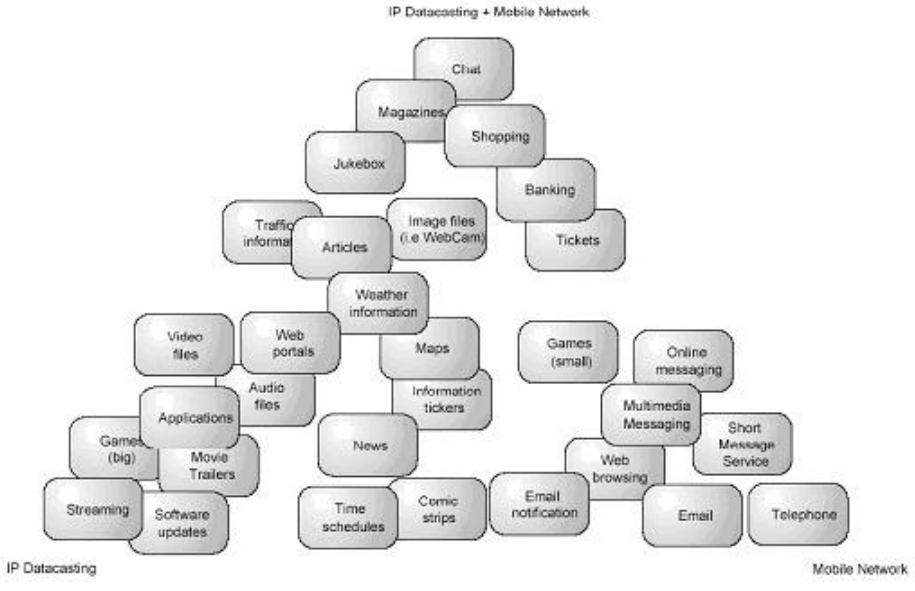


Abbildung 2.3: Übersicht möglicher IPDC basierter Dienste. [11]

**Fazit**

An den Beispielen zeigt sich, dass die Möglichkeiten von Übertragungsdiensten mittels DVB den anderen Technologien sehr ähnlich sind, was natürlich zu einer Konkurrenzsituation führt. Zusätzlich muss man mit einbeziehen, dass all diese DVB-H Services bisher nur auf dem Papier bestehen, erst in Skandinavien werden kleinere Feldversuche durchgeführt. Wirklich durchsetzen werden sie sich wohl erst, wenn sich DVB im TV-Bereich schon etabliert hat. Da gerade die Interaktiven Dienste wegen des Antwortkanals auf schon bestehenden Technologien aufbauen, ist DVB-H auch abhängig von Konkurrenztechnologien wie beispielsweise UMTS - je nach Dienst unterschiedlich stark [11].

### 2.2.2 TV

Das Hauptaugenmerk bei der Entwicklung von DVB liegt jedoch nicht bei den IP-Services, sondern bei der Übertragung von digitalem Fernsehrundfunk. DVB soll dank Qualitäts- und anderen technischen Vorteilen das analoge Fernsehen in Zukunft ganz ersetzen. Gerade aber weil das analoge Fernsehen stark verbreitet ist und sich als Standard längst hat festsetzen können, wird der Übergang von der jetzigen Situation zum gewünschten digitalen Zeitalter wohl noch Zeit brauchen. Angesichts der starken Förderung dieses europäischen Standards, ist an einer zukünftigen Einführung jedoch kaum noch zu zweifeln [11].

**Geschichte**

Die erste Verbreitung von digitalen TV-Angeboten fand im Bereich PayTV statt. Im deutschsprachigen Raum startete 1996 der digitale Sender DF1 - damals noch Teil des Kirch Media Imperiums - sein digitales Angebot gegen eine monatliche Gebühr auszustrahlen [29]. Die Versprechen waren gross: Anschauen was man will, wann man es will, bei besserer Qualität von Bild und Ton und riesiger Auswahl; all das gegen eine moderate monatliche Gebühr. Das Konzept der Themensender wurde eingeführt, von Action über Sport und Kindersendungen bis zu Dokumentationen und Filme für Erwachsene war alles bestellbar. Für den Empfang war ein digitaler Receiver notwendig, der auch heute noch benötigt wird. Trotz des verlockenden Angebotes und der technischen Qualitätsverbesserung sind die Anteile an Digitalkunden im Vergleich zum analogen Fernsehen immer noch sehr gering. DF1 wurde nach wenigen Jahren in Premiere umgetauft, das Programm wurde noch stärker diversifiziert. Hier in der Schweiz wird das Premiere Angebot von Teleclub vertrieben. Auch Teleclub wirbt momentan stark auf ihrer Webseite dafür, dass ihre Kunden nun endlich von analog auf digital umstellen sollten, da sowohl Qualität wie auch Angebot besser sei. Die Übertragung erfolgt über Satellit und Kabel. Doch nicht nur Pay-TV ist verfügbar, auch Sender wie ARD, ZDF, RTL und CNN und verschiedene andere private Sender bieten unterdessen längst digitale Sendungen, die neben dem bekannten analogen Programm auch Inhalte bieten, die nur digital empfangbar sind [12].

### Verbreitung

Im Juni 2003 hat der Bundesrat persönlich der SRG die Erlaubnis erteilt, digital zu senden [13]. In der Schweiz ist jedoch nicht die Übertragung via Satellit oder Kabel als Ziel angestrebt, sondern es wird versucht, ein umfassendes, terrestrisches Rundfunknetz basierend auf dem DVB-T Standard zu realisieren. Durch den Eintritt der SRG wurden die ersten gewichtigen Voraussetzungen geschaffen, dass DVB-T definitiv auch bei uns eingeführt werden kann. Die SRG durfte von da an nun per Konzessionsentscheid ein erstes Sendernetz für die Verbreitung von vier eigenen Programmen realisieren; Privatsender waren noch nicht beteiligt. Technisch sind in der Schweiz insgesamt 6 DVB-T-Ketten möglich, à je 4 Sender. Ganz im Sinne des Service Public bestand das Angebot hier aus den zwei deutschschweizerischen Sendern - SF1 und SF2 - und je einem aus den andern zwei Sprachregionen. In den andern Sprachregionen sah das Angebot entsprechend vertauscht aus. Die SRG wollte übrigens im Zuge der Technologisierung die Gebühren erhöhen, wegen der Mehrkosten die eine zeitgleiche digitale Ausstrahlung so mit sich zieht. Zwar wurde ihr Anliegen abgeschmettert, aber es ist dennoch mit einer Erhöhung der Gebühren zu rechnen, wenn sich das DVB-Netz auf die ganze Schweiz ausdehnen wird [13] [14] [15].

Begonnen wurde mit einer Testausstrahlung im Engadin. Dank den erzielten Erfolgen - trotz mittelmässiger Euphorie seitens der Bevölkerung - erfolgte eine Gebietsausdehnung in das Tessin und in die Romandie. Bis ins Jahre 2009 sollt laut Auftrag der Regierung die ganze Schweiz erschlossen sein; 2020 muss die Ablösung des analogen TV erfolgen. In Deutschland existieren bereits mehrere Angebote an DVB-S und DVB-C. Erst in kleineren Ausbreitungszonen steht auch DVB-T zur Verfügung. Auf lange Sicht wird aber auch in Deutschland eine komplette Abdeckung des ganzen Landes angestrebt [16].

### Fazit

DVB hätte für die Schweiz mehrere Vorteile. Die Technologie würde es ermöglichen, in der gesamten Schweiz auf der gleichen Frequenz zu senden, ohne grosse Interferenzen befürchten zu müssen. Die Bild- und Tonqualität würde ansteigen, Sender wie Pro7 und Sat1 senden all ihre digitalen Übertragungen in Dolbi digital 5.1 aus. Ausserdem wäre es ein politisches Ziel über ein eigenes, von ausländischen Kabelanbietern unabhängiges digitales Ausstrahlungsnetz verfügen zu können, das die ganze Schweiz versorgen kann. Doch obwohl DVB von vielen Seiten stark gefördert wird und die Qualität eigentlich besser als die vom herkömmlichen Fernsehen ist, setzt sich DVB nur langsam durch. Der grösste Teil aller Leute die eine solche Set-Top-Box besitzen, haben diese nur wegen Premiere oder anderen digitalen Anbietern. Von diesen Pay-TV Angeboten abgesehen, sieht die DVB-Landschaft aber noch karg aus. Die Mehrheit der Leute hat bisher keinerlei Grund zu wechseln, viele wissen nicht einmal, dass DVB existiert. Die meisten IP-Services bestehen nur auf dem Papier, und bereits jetzt kursieren Wirtschaftliche Untersuchungen, die zu zeigen versuchen, dass sich DVB nicht durchsetzen wird.

Nicht nur das mittelmässige Angebot ist jedoch das Problem, es mangelt mitunter auch an der Nachfrage. Denn wieso sollte ein TV-Konsument momentan auf DVB umsteigen? Wer besseres Programm will, der holt sich PayTV - für digitales Schweizer Fernsehen werden

wohl die wenigsten extra eine Set-Top-Box anschaffen - bessere Bildqualität hin oder her. Für den normalen TV-Konsumenten sind die Einstiegskosten im Vergleich zur Leistung noch zu hoch; analoges Fernsehen genügt. Die Zukunft von DVB-H und den IP Diensten ist noch unklarer - tatsächlich genutzte Angebote innerhalb der nächsten fünf Jahren würden überraschen. Solche Angebote werden wohl erst auftauchen, wenn sich DVB im TV-Sektor schon etablieren konnte. Doch auch davon ist man noch weit entfernt, speziell auch in der Schweiz. Damit DVB auch bald wirklich zu dem führenden Standard werden kann, muss speziell auch die Verbreitung im öffentlichen-rechtlichen Rundfunksektor seitens der SRG vorangetrieben werden [17].

Eine Senkung der Einstiegskosten ist relativ unwahrscheinlich, deshalb muss der Mehrwert für die Kunden über eine Verbesserung des Angebotes herbeigeführt werden - des weiteren würde es nicht Schaden, wenn der Bevölkerung der Standard näher gebracht würde. Ansonsten wäre wohl die einzig mögliche Lösung die Zwangsumsiedlung vom analogen auf das digitale Fernsehen. Wenn einfach nur noch digital gesendet wird, erleidet das analoge Fernsehen automatisch den endgültigen Todesstoss. Digitales Fernsehen ist definitiv unser Fernsehen der unmittelbaren Zukunft, nur bleibt die Frage wie lange es noch bis zur Einführung dauern wird.

Die Prognosen scheinen optimistisch, gerade weil der Ausbau noch nicht zügig vorankommt. Das Projekt DVB wird jedoch kaum scheitern, da wichtige Institutionen den Wert erkannt haben, und die Verbreitung auch auf der politischen Ebene als Ziel gesetzt wurde. DVB wird kommen, früher oder später. Spätestens wenn sich DVB international durchgesetzt hat, muss die Schweiz mitziehen.

### Digital TV am PC

Wer seine Freizeit lieber vor dem PC als vor dem Fernseher verbringt, der sollte es sich überlegen, anstatt einer analogen TV-Karte direkt auf digitales Fernsehen zu setzen. Ob für Laptop oder PC, DVB bietet für beide Lösungen komfortablen TV-Empfang. Komfortfunktionen wie zeitversetztes Fernsehen oder gar voll digitales Videorecording lassen sich viel leichter realisieren als bei analog empfangenem Material. Für den Empfang digitaler Programme über den PC wird eine DVB-Karte benötigt, die über einen USB Anschluss mit dem PC gekoppelt werden kann. Eine solche DVB-Karte ist heutzutage nicht einmal mehr teurer als eine gleichwertige analoge TV-Karte für ungefähr 150 Franken ist man im Rennen. Die Bildqualität dagegen ist wie auch bei den andern DVB-Varianten besser, und da die digitalen Programme im MPEG-2-Format gesendet werden, kann das Fernsehprogramm auch direkt auf der PC-Festplatte abgespeichert werden. Ebenfalls möglich ist der Empfang verschlüsselter PayTV Programme sowie digitaler Zusatzdienste, Internet inklusive.

Die grösste Verbreitung von digitalem Rundfunk- und Fernsehprogrammen erfolgt im speziellen in Deutschland bisher über Satellit und so verwundert es nicht, dass für den Satellitenempfang entsprechend auch die grösste Auswahl an TV-Karten für den PC existiert. Aber auch DVB-T Varianten werden Angeboten, pünktlich zum Start des terrestrischen Rundfunks. Für die Aufnahme selber genügt ein PC, der vor vier Jahren mal neu war.

Die Wiedergabe setzt jedoch voraus, dass der MPEG-2 codierte Datenstrom wieder decodiert werden kann. Günstigere DVB-Karten erledigen das jedoch nicht direkt mit ihrer eigenen Hardware, sondern benötigen eine zusätzliche Software. So unterscheiden sich die meisten Angebote auch nicht in der Hardware, sondern in der mitgelieferten Software - jedoch erledigt nicht jede Software ihre Aufgabe gleich gut. Für die Decodierung muss der PC die entsprechende Rechenleistung zu Verfügung stellen können. Auf Maschinen mit weniger als 1 GHz kann die Wiedergabe nur mit Mühe und Qualitätseinbussen abgespielt werden. Das Abspielen selber benötigt kaum irgendwelche Festplattenkapazität, das Speichern jedoch umso mehr. Für 90 Minuten Film muss mit einer Belegung von bis zu 10 GB Speicher gerechnet werden. Entweder kauft man sich entsprechende Harddiskkapazitäten oder einen DVD-Brenner und einen grossen Stapel DVD-Rohlinge [12].

### Archivierung auf DVD

Um Mitschnitte aus dem Digital-TV verlustfrei zu schneiden und so herzurichten, dass sie auch von einem herkömmlichen DVD-Player abgespielt werden können, müssen zuerst verschiedene Hürden genommen werden. Denn obwohl sowohl DVD wie auch DVB beide auf dem MPEG-2 Format basieren, sind sie nicht automatisch miteinander kompatibel. Feine Unterschiede führen dazu, dass eine direkte Wiedergabe unmöglich ist. Mögliche Probleme die auftreten können sind: zu hohe Bitraten, falsche Auflösungen und zu lange Groups of Pictures[1].

Mit Hilfe verschiedener Tools lassen sich diese Probleme aber relativ einfach umgehen. Zu allererst müssen Audio- und Videospuren voneinander getrennt werden, was erst eine weitere Bearbeitung ermöglicht. Anschliessend kann die gewünschte Auflösung modifiziert werden, und das Format geändert werden. Um die lästigen Werbeunterbrechungen aus dem Programm zu schneiden, wähle man ein Schneidprogramm nach Wahl. Ohne Werbung werden sowohl Nerven wie auch Speicherkapazität geschont, und die Lieblingssendungen können ohne Unterbruch genossen werden. Zuletzt bannt man die entstandenen Videodaten noch auf DVD, und kreiert ein dazu passendes Menü um das Sehvergnügen noch zu vergrössern. All die für diesen Prozess benötigte Software lässt sich kostenlos und legal beschaffen. Wer jedoch irgendwelche Softwarepräferenzen hat, kann diese getrost nutzen.

Die Archivierung der TV-Übertragungen auf digitalen Medien hat verschiedene Vorteile. Zum einen spart man Festplattenspeicher, der in Massen nötig ist für die Speicherung der Übertragungen, zum anderen ermöglicht man eine Platz sparende Archivierung in digitaler Bildqualität - Schluss mit Videostapeln mit verzerrten Aufnahmen. Für viele die auf digitales Fernsehen via PC umsteigen werden, wird die saubere Auslagerung des Programms auf DVDs sicher eine interessante Option darstellen. Für einige wäre dies sicher auch das Killerargument sich einen solchen Empfänger anzuschaffen [18].

[1] Als Group of Pictures wird die Gruppe von aufeinanderfolgenden Bildern innerhalb zweier Vollbilder eines mittels MPEG-2 codierten Videos bezeichnet. Zwischen allen Bildern einer Gruppe herrschen Abhängigkeiten, die bei der Codierung ausgenutzt werden. Die Effizienz der MPEG2-Kompression beruht auf diesen Groups of Pictures.

Die gesetzlichen Schranken sind die selben wie auch im analogen Bereich: Solange die erstellte Kopie nur für den privaten Gebrauch verwendet wird, spielt sich alles im legalen Rahmen ab. Die digitale Kopie darf genau wie das analoge Videoband an Freunde und Bekannt ausgeliehen werden, und von denen für den privaten Gebrauch kopiert werden. Es ist jedoch nicht gestattet die digitale Kopie - beispielsweise via Internet - einem grossen Publikum zugänglich zu machen, ohne die offizielle Zustimmung des Urhebers zu besitzen [37].

## 2.3 Technik

Dieses Kapitel führt in die technischen Grundlagen des Rundfunks ein und beschreibt die Standards des digitalen Rundfunks. Weiterführend werden Konzepte für das IP basierte Senden von Daten in bidirektionalen mobilen Breitbandnetzwerken vorgestellt. Nachdem im ersten Teil schon einige Rundfunkbegriffe beiläufig erwähnt wurden, werden sie nachfolgend als Einstieg in den technischen Teil präzise definiert.

### 2.3.1 Eine kleine Rundfunkterminologie

Als Radiowellen werden elektromagnetische Wellen in einem Frequenzbereich bezeichnet, der für die drahtlose Übertragung von Sprache, Bildern und anderen Daten verwendet wird. Z.B via Rundfunk oder mittels Satellit. Radiowellen haben einen Frequenzbereich von rund 75 kHz bis 10 GHz. Dies entspricht einem Wellenlängenbereich von rund 10 Km bis 1 cm. Die Wellen mit Längen im Kilometerbereich werden für Radiorundfunk verwendet. Im Meterbereich werden Fernsehsignale übertragen und die Wellenlängen von unter 1 Meter sind für den Mobilfunk und das Satellitenfernsehen vorgesehen [19]. Tabelle 2.2 gibt eine Übersicht. Elektromagnetische Wellen bewegen sich je nach Umgebung nahezu mit Lichtgeschwindigkeit fort.

Tabelle 2.2: Übersicht Radiowellen.

| Bezeichnung | Frequenzbereich | Verwendung |
|---|---|---|
| LW, MW, KW [2] | 30 kHZ -30 MHz | Radio |
| UKW | 30 - 300 MHz | Radio, Antennenfernsehen |
| Dezimeterwellen | 300 MHz - 3 GHz | Antennenfernsehen, Mobilfunk |
| Zentimeterwellen | 3 - 30 GHz | Satelliten TV |

Unidirektionale Informationsübertragung mit Hilfe von elektromagnetischen Wellen als Trägermedium in Bezug auf die Empfängergruppe kann in drei Modi geschehen:

- Broadcast: Eine Sendestation sendet Informationen, die für alle Empfänger innerhalb des Senderadius bestimmt und uneingeschränkt empfangbar sind. Z.B. öffentliches Radio.

[2] Lang-, Mittel- und Kurzwellen.

- Multicast: Eine Sendestation sendet Informationen, die nur für eine Teilmenge aller Empfänger innerhalb des Senderadius bestimmt sind. Z.B. gebührenpflichtiger Content für mobile Geräte. Die Nutzungseinschränkung kann mit Datenverschlüsselung oder Conditional Access erzwungen werden.
- Unicast: Eine Sendestation sendet Informationen, die nur für einen Empfänger innerhalb des Senderradius bestimmt sind (Point to Point). Z.B. Mobilfunkgespräch.

Um Information auf einer sinusförmigen elektromagnetischen Welle zu übertragen, muss diese so beeinflusst werden, dass die zu übertragende Information darin codiert ist. Die Sinuswelle kann auf drei Arten verändert werden:

- Amplitude
- Frequenz
- Phase (Verschiebung)

Das Verändern von elektromagnetischen Wellen zur Informationsübertragung wird Modulation genannt. Amplitudenmodulation (AM) und Frequenzmodulation (FM) für die Übertragung von Radioprogrammen sind die zwei bekanntesten Verfahren.

Um genügend Informationen übertragen zu können, die ein Radio- oder Fernsehprogramm darstellen, braucht es mehrere unterschiedlich lange elektromagnetische Wellen. Die für die Übertragung eines Programmes verwendeten Wellenlängen, respektive die Frequenzen, werden aus praktischen Gründen aufeinanderfolgend gewählt. Dieses Bündel von Trägersignalen wird als Bandbreite oder Frequenzband bezeichnet. Genauer definiert bezeichnet die Bandbreite die Differenz zwischen der grössten und der kleinsten Frequenz in einem kontinuierlich zusammenhängenden Frequenzbereich: Bandbreite B = (Frequenz maximal - Frequenz minimal) [20]. Das geometrische Mittel zwischen diesen Frequenzen ist die Mittenfrequenz, die umgangsprachlich ein Frequenzband bezeichnet. Z.B ist Radio DRS 3 auf 89.55 MHz in der Region Grenchen zu empfangen.

Für die Verbreitung eines analogen Fernsehkanals via terrestrische Sendestation ist eine Bandbreite von rund 8 MHz notwendig. Aufgrund der Kompression bei digital vorliegenden Daten und den digitalen Codierungsverfahren können pro 8 MHz Band je nach Qualität typischerweise 4 bis 5 digitale Fernsehkanäle übertragen werden. Die verschiedenen Modulationsverfahren für analoge und digitale Informationen sind vom Prinzip her ähnlich: Die Trägersignale werden in Amplitude, Frequenz und Phase so verändert, dass sie Informationen repräsentieren. Im digitalen Rundfunk ist es zusätzlich möglich, mehrere Programme gebündelt zu versenden. Das Bündeln der Programme nennt sich Multiplexing.

### 2.3.2 Die MPEG Standards

Dieser Abschnitt gibt eine kurze Übersicht zu den MPEG Standards. Nachfolgend wird detaillierter auf den MPEG-2 Standard eingegangen, weil dieser hauptsächlich in DVB

zur Anwendung kommt. Die Moving Pictures Expert Group (MPEG) wurde 1988 von der ISO mit dem Ziel Standards für die Kodierung von digitalen Audio und Videodaten zu entwickeln gegründet. Die Gruppe besteht heute aus rund 350 Mitgliedern: Firmen aus verschiedenen Industrien sowie Hochschulen [23].

Die MPEG Standards sind so ausgelegt, dass die Anwender (z.B. Hersteller von Decodern) MPEG Applikationen nach ihren Bedürfnissen selber ausgestalten können , u.a. Bildqualität oder Hardwarekosten von Encoder Chips. Die Encoder- und Decoderprodukte auf dem Markt sind entsprechend vielfältig. MPEG Codecs[3] wenden verlustbehaftete Kompressions- bzw. Dekompressionsverfahren an. Bei der Dekompression sind die Verluste allerdings so gering, dass der Output der Dekompression dem Originalinput genügend ähnlich ist. Kodierungssysteme für bewegte Bilder, wie MPEG-2, wenden Verfahren an, um den Inhalt eines Bildes aufgrund von vorhergegangenen Bildern zu bestimmen. Es werden dann nur die Differenzen zu den vorhergegangenen Bildern kodiert, was zu grossen Dateneinsparungen führt. Der Videoteil von MPEG-2 unterstützt Interlaced Video, wie es in den analogen TV Standards PAL und NTSC angewendet wird. Beim Interlacing werden auf ein rastergescanntes Display alternierend einmal die geraden Linien, und darauf folgend einmal die ungeraden Linien projiziert. Der Effekt dieser Technik ist ein verringertes Flimmern des Bildes. Die B Tabelle 2.3 gibt eine grobe Übersicht zu den Standards.

Tabelle 2.3: Übersicht der MPEG Standards.

| Standard | Funktion | Anwendungen |
|---|---|---|
| MPEG-1 | Video- und Audiokompression | MP3, Video CD |
| MPEG-2 | Interleaced Video- und Audiokompression | DVD, DVB |
| MPEG-3 | Aufgegeben [4] | |
| MPEG-4 | MPEG-1 und -2, VRML, Rights Management | Videophone, Streaming |
| MPEG-7 | Multimedia Content Description | Streaming (mit MPEG-4) |
| MPEG-21 | Multimedia Framework | zukünftiger Standard |

### Die Funktionsweise von MPEG-2

Zu Beginn des DVB Projektes am Anfang der Neunzigerjahre wurde MPEG-2 als Standard zur Übertragung von Daten gewählt. MPEG-2 entsprach den Anforderungen für den ersten herausgegebenen DVB Standard, DVB-S. Ein 33 MHz Satellitenband kann typischerweise ca. 52 MBit/s übertragen und MPEG-2 ist in der Lage solche Kompressionsraten zu erreichen [24]. Zudem ist MPEG-2 ein offener und erweiterungsfähiger Standard, genau wie es die DVB Standards werden sollten. Es gab somit keinen Grund ein eigenes Kompressionsformat zu entwickeln. Der MPEG-2 Standard besteht zur Zeit aus neun Teilen. Die ersten drei Teile sind international standardisiert, weitere Teile sind unterschiedlich weit fortgeschritten in der Entwicklung. Die MPEG-2 Teilstandards geben ein Framework vor wie die einzelnen Teile (Audiospuren, Videospuren, Metadaten und persönliche Benutzerdaten) eines Programmes (z.B. analoge Daten, DVD) zu formatieren

[3] Compressor-Decompressor, Hardware oder Software zum Komprimieren/Dekomprimieren von Daten.
[4] Ursprünglich für HDTV entwickelt, MPEG-2 erwies sich als genügend.

und zu dekomprimieren sind. MPEG-2 legt auch fest wie die Komponenten[5] zu einem synchronen Bitstream zusammengefügt werden (Multiplexing) als Vorstufe für die effiziente Übermittlung (z.B. via ADSL oder DVT-T). Die Codierung der Informationen ist hingegen den Implementierern von MPEG-2 überlassen solange das Datenformat eingehalten wird - als Voraussetzung für eine Dekodierung bei einem beliebigen MPEG-2 kompatiblen Endgerät. Es haben sich eine Reihe von Algorithmen für die Codierung etabliert. Meistens werden die Varianten des Discrete Cosine Transform (DCT) verwendet, ein auf Fourrier Transformationen basierendes Verfahren [30]. Anhand von Bild 2.4 wird die Anwendung von MPEG-2 verdeutlicht.

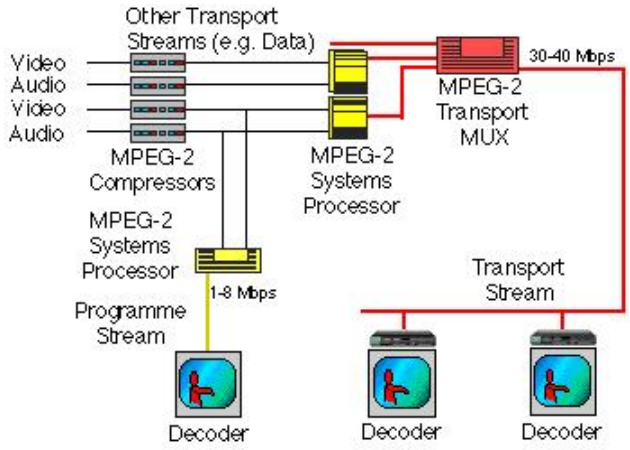


Abbildung 2.4: Schematische Darstellung des MPEG-2 Systems. [41]

Rohdaten, z.B. eine analoge Fernsehsendung, müssen zuerst mit MPEG-2 Kompressoren kodiert werden (Encoding). Der Encoder kann ein Hardwarechip oder softwarebasiert sein. MPEG-2 Encoder haben eine eigene Logik, oft werden proprietäre Algorithmen angewendet. Video- und Audioinformationen werden separat komprimiert. Die komprimierten Outputstreams der Encoder werden Elementary Streams (ES) genannt. Mögliche ES Arten sind:

- Kontrollstream
- Audiostream
- Videostream
- Metadatenstream

Als nächstes wandert jeder ES in einen MPEG-2 Systemprozessor, der die einzelnen Streams zusammenführt und sie in Pakete bündelt. Die Pakete werden mit einem 6 Byte Protokoll Header versehen. Die Pakete können bis zu 65536 Bytes lang sein. Der Header enthält Informationen über die Länge des Pakets sowie eine ID, die das Paket als einen Typ (Video, Audio, Control) signiert. Die Outputpakete aus dem MPEG-2 Prozessor stellen

[5] Als Komponente wird hier ein verarbeiteter Teil eines Programmes verstanden, z.B. die komprimierten und formatierten Videodaten.

vollständige Streams dar (z.B. eine Videosequenz mit Ton), allerdings in Pakete aufgeteilt (Packetised ES (PES)). Im Falle einer Übertragung mittels eines Trägermediums des Streams werden die PES in 188 Byte grosse Pakete aufgeteilt, 4 Byte sind im Header für die Identifizierung des Paketes reserviert. Die sich im Header befindende Packet Identifier (PID) Nummer legt fest zu welchem Stream (Audio, Video) innerhalb des PES das 188 Byte grosse Paket gehört. Bild 2.5 stellt den Sachverhalt dar. Für die physikalische Übertragung der 188 Byte grossen Pakete spezifizieren die DVB Standards unterschiedliche digitale Modulationsverfahren in Abhängigkeit vom Übertragungsweg [21].


Abbildung 2.5: Ein MPEG Transportstream für ein TV Programm. [41]

### 2.3.3 Technische Übersicht der DVB Varianten

Die DVB Projektgruppe entwickelt und verfasst Standards, welche dann vom ETSI herausgegeben werden. DVB umfasst eine Menge von Standards die ein System[6] für digitales Fernsehen spezifizieren. Die Standards geben Vorgaben wie sämtliche anfallenden Datentypen: Video, Audio und komplementäre Informationen (z.B. Electronic Program Guide) codiert, formatiert, übertragen und letztlich decodiert werden. Allerdings sind die DVB-Produkte Hersteller frei im Gebrauch von MPEG-Kompressionsalgorithmen, solange der Output das DVB Datenformat respektiert. Video- und Audioinformationen werden derart mit MPEG-2 komprimiert, dass sich eine bessere als die gewohnte PAL/SECAM Bildqualität ergibt bei wesentlich geringeren Datenraten [22]. Die Standardisierung und Anwendung der vier DVB Übertragungsvarianten sind unterschiedlich weit fortgeschritten. Die schematische Übersicht 2.6 illustriert die DVB-Varianten.

Flexibilität war ein Hauptdesignziel für DVB. Das Transportsystem von DVB ist generisch, es gibt keine Restriktionen bezüglich der Art der gesendeten Daten. Es wurden Übertragungstechniken (Transportstreams) definiert für folgende Datentypen:

- Audio
- Video
- Elektronische Programmführer (EPG, Multimediastil)
- Elektronische Serviceführer (ESG, Teletextstil)
- Pay-per-view TV

[6] Es wird ein Baseline System spezifiziert, d.h. ein minimales System für das Erfüllen der Anforderungen.

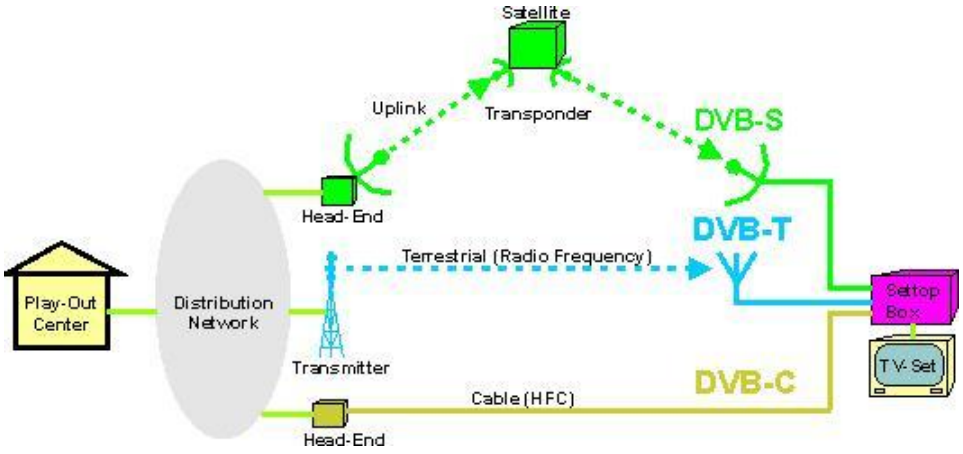


Abbildung 2.6: Die Übertragungswege der DVB Typen. [22]

- Datenkarusselle (periodisches Senden von Informationen)
- IP Paketübertragung

Die einzelnen Standards finden unterschiedlich Anklang bei den Systemimplementierern. Der elektronische Programmführer wird beispielsweise kaum genutzt. Eine wichtige Voraussetzung für interaktive mobile Systeme sind die letzten beiden gelisteten Übertragungstechniken: Datenkarusselle und IP Paketübertragung. Auch kann die Nutzung eines Frequenzbandes (typischerweise 8 MHz) nach den Bedürfnissen des Anwenders ausgestaltet werden. Es können beliebige Streams für die Übertragung gemultiplext werden. So finden z.B. in einen Übertragungsstream Fernseh- und Radioprogramme sowie IP-Daten Platz.

### 2.3.4 DVB-T Modulations- und Übertragungsverfahren

"Für DVB-T sind drei Modulationsverfahren festgelegt worden: 4-QAM, 16-QAM und 64-QAM. Sie erfüllen gemeinsam mit weiteren wählbaren Systemparametern unterschiedliche Anforderungen an Übertragung und Empfang. So kann z.B. der Schutz gegen Übertragungsfehler unterschiedlich hoch gewählt werden. Durch diese Flexibilität besteht die Möglichkeit, die Anzahl der auszustrahlenden Programme und die Empfangsart zu bestimmen, also z.B., ob mobil, portabel mit Stabantenne oder stationär empfangen werden soll" [26].

Bei der digitalen Quadraturamplitudenmodulation (QAM) werden die Modulation der Amplitude und der Phase kombiniert. Ein Generator erzeugt zunächst zwei Sinusschwingungen mit gleicher Frequenz und einem Phasenunterschied von 90°. Diese beiden Trägerschwingungen durchlaufen dann jeweils einen Amplitudenmodulator. Die Amplitude wird je nach angewendetem QAM Verfahren verändert:

- 4-QAM: Jeder der beiden Träger kann zur Darstellung von einem Bit (2 Stufen) mit den Faktoren -1 und 1 multipliziert werden.

- 16-QAM: Jeder der beiden Träger kann zur Darstellung von 2 Bits (4 Stufen) mit den Faktoren -3, -1, 1 und 3 multipliziert werden.
- 64-QAM: Jeder der beiden Träger kann zur Darstellung von 3 Bits (8 Stufen) mit den Faktoren -7, -5, -3, -1, 1, 3, 5 und 7 multipliziert werden.

Nach der Amplitudenmodulation werden die beiden modulierten immer noch um 90° phasenverschobenen Trägerschwingungen addiert und sind dann bereit für die Übertragung. Die Summe der beiden Trägerschwingungen kann in der planaren Ebene als Vektor mit zwei Komponenten interpretiert werden. Abhängig von den gewählten Multiplikationsfaktoren sind dabei die Länge (Amplitude) und der Winkel (Phase) dieses Vektors einstellbar. Man kann den Vektor damit auf verschiedene Punkte zeigen lassen und auf diese Art und Weise Daten übertragen [27] [28]. Die Zahlen 4, 16 und 64 im Zusammenhang mit QAM beziehen sich auf die Anzahl der darstellbaren Punkte mit dem Zweikomponentenvektor. Je mehr Punkte darstellt werden, desto unklarer sind sie voneinander zu unterscheiden. Die Wahrscheinlichkeit von Fehlern bei der Decodierung aufgrund von Störsignalen bei der Übertragung nimmt somit bei vielen mittels Vektor dargestellten Punkten zu. 4-QAM ist somit das robuste Übertragungsverfahren. Da bei 16 und 64-QAM mehr Fehler auftreten, können zusätzlich Fehlerkorrekturverfahren eingesetzt werden mit dem Preis einer geringeren Nettodatenübertragungsrate. Tabelle 2.4 zeigt die Datenraten der QAM Modulationsverfahren bei idealen Bedingungen. Jedem Punkt lässt sich ein Symbol zuordnen, das z.B. eine Zahl oder ein Zeichen aus einem Zeichensatz repräsentiert. n-QAM bedeutet, dass mit der Modulation einer Periodenlänge des Trägersignals eines aus n Symbolen ausgesendet werden kann [28].

Tabelle 2.4: Die QAM Modulationsverfahren mit den maximal möglichen Datenübertragungsraten.

| Verfahren | Datenrate (Mb/sec) |
|---|---|
| 4-QAM | 10.6 |
| 16-QAM | 21 |
| 64-QAM | 30 |

Da mit z.B. mit 64-QAM nur 6 Bit (26 = 64) mit Hilfe einer Trägerfrequenz (zwei modulierte und aufsummierte Trägerschwingungen) fortlaufend übertragen werden können, müssen mehrere Frequenzen mit 64 QAM moduliert werden für das Senden von Audio- oder Videoinformationen. Im Falle von DVB-T wurden zwei Bänder festgelegt, die man auch als Modi bezeichnet [22]:

- 2k Modus: 2048 mögliche Trägerfrequenzen, effektiv 1704 verwendet.
- 8k Modus: 8192 mögliche Trägerfrequenzen, effektiv 6817 verwendet.

Dieses Verteilen von Informationen auf viele, dicht nebeneinander liegende Trägerfrequenzen mittels QAM Modulation wird auch Coded Orthogonal Frequency Division Multiplex Verfahren (COFDM) bezeichnet. Man spricht in diesem Zusammenhang auch von Mehrträgerverfahren. COFDM ist eine Erweiterung von Orthogonal Frequency Division

Multiplexing (OFDM). COFDM bietet zusätzlich zu OFDM einen Fehlerschutz um hohe Stabilität beim Empfang zu gewährleisten [31]. Der 2k Modus eignet sich gut für den mobilen Einsatz, da Empfang theoretisch bis rund 800 km/h möglich ist. Allerdings darf der Abstand zum Sender nicht grösser als 17 km sein. Im 8k Modus kann nur bis zu Geschwindigkeiten von 200 km/h empfangen werden, dafür ist der maximale Empfangsradius 68 km [22].

Beim im nächsten Abschnitt beschriebenen DVB-H Standard wurde ein Kompromiss gemacht zwischen Empfangsgeschwindigkeit und Senderadius, indem zusätzlich ein 4k Modus eingeführt wurde. Die oben genannten Geschwindigkeiten sind allerdings mit Vorsicht zu geniessen, im Internet finden sich verschiedenste Geschwindigkeitsangaben in Erfahrungsberichten. Mit DVB-T können im Gegensatz zu analogem Rundfunk, auf einem Kanal (Band) mehrere Sendeanlagen an verschiedenen Standorten im sogenannten Gleichwellenbetrieb (Single Frequency Network, SFN) für die Ausstrahlung eines identischen Transportstroms betrieben werden. Damit wird im Vergleich zum heutigen analogen Sendebetrieb die Effektivität der Frequenznutzung erhöht. Der Gleichwellenbetrieb bietet außerdem verbesserte Empfangsbedingungen im Versorgungsgebiet durch den entstehenden Netzgewinn [38].

### 2.3.5 Weiterentwicklung von DVB-T: DVB-H

Ursprünglich war DVB-T nicht vorgesehen für den mobilen Empfang, suboptimal war es dennoch möglich. Im Jahre 2002 wurde eine Gruppe gebildet um einen neuen DVB Standard für mobile Geräte zu entwickeln abgeleitet von DVB-T. Der neue Standard wird als DVB-X, DVB-M (DVB-Mobile) oder DVB-H (DVB-Handheld) bezeichnet. Berücksichtigt werden die Anforderungen rund um den mobilen Empfang: Baterielaufzeit bei den Empfängergeräten, verbesserte Fehlerkorrektur und optimierter Empfang wenn das Empfangsgerät sich in Bewegung befindet. Die angestrebten technischen Zielparameter sind 15 Mbit/s in einem 8 Mhz Band im Gleichwellenbetrieb. Die Ziele sollen erreicht werden ohne die Kompatibilität mit existierenden DVB-T Netzen zu gefährden. Der DVB-H Standard wurde im November 2004 vom ETSI anerkannt.

DVB-H unterscheidet sich von DVB-T hauptsächlich durch den Time-Slicing Mechanismus. Dabei werden die Daten nicht wie bei DVB-T kontinuierlich zum Empfänger gesendet, sondern schubweise (Burst) und dann lokal abgespeichert. Das Endgerät muss dadurch nicht permanent auf Empfang sein, was bis zu 90% weniger Stromverbrauch führt. Neu wurde auch ein 4k Modus mit 3409 aktiven Trägerfrequenzen spezifiziert, um einen auf mobile Bedürfnisse zugeschnittenen Kompromiss zwischen Reichweite und Datendurchsatz zu erhalten. Es wurden Verbesserungen vorgenommen um Interferenzen zu verringern und die Fehlerkorrektur erweitert um den generell kleinen mobilen Empfängerantennen gerecht zu werden [39]. DVB-H verwendet wie DVB-T die Quadraturamplitudenmodulation. Neu ist die Unterstützung von MPEG-4 als Kompressionsformat, ausgelegt für Zielapplikationen von DVB-H: Streaming und Videophone. MPEG-4 unterstützt auch Digital Rights Management. DVB-H ist nicht grundlegend neu, sondern DVB-T wurde auf mobilen Empfang optimiert. Die neueren technischen Entwicklungen wurden aufgegriffen um eine Grundlage für die mobilen Benutzerbedürfnisse der Zukunft zu schaffen.

### 2.3.6 DVB-S

DVB-S spezifiziert die Übertragung von DVB Inhalten per Satellit. DVB-S wendet das Modulationsverfahren der Quadraturphasenumtastung (Quadrature Phase Shift Keying, QPSK) mit hohen Fehlerkorrekturanteilen an. QPSK ist vom Konzept her identisch zu dem im Zusammenhang mit DVB-T besprochenen 4-QAM Modulationsverfahren. Ein typischer Satellitenkanal hat 36 MHz Bandbreite und ermöglicht Übertragungsraten bis rund 40 Mbps. Da bei Satellitenverbindungen grosse Bitfehlerraten auftreten, werden die Transportpakete mit Prüfsummen gesichert: Zu jedem 188 Byte grossen Transportpaket werden 16 Prüfbytes zugefügt um Forward Error Correction zu ermöglichen. Bei der Forward Error Correction hat der Empfänger bis zu einem gewissen Grad die Möglichkeit bei der Übertragung aufgetretene Fehler aufgrund der beigefügten Redundanz zu korrigieren. Bei dem insgesamt 204 Bytes langen Transportpaket können bis zu 8 korrupte Bytes repariert werden [22]. Ein Satellit ist mit mehreren Transpondern (Sender) ausgestattet. Beim Senden via einen 38 Mbps Transponder mit DVB-S sind folgende typischen Nutzungsszenarien denkbar:

- 4-8 TV Kanäle
- 150 Radioprogramme
- 550 64 kbps Datenkanäle

"Bei DVB-S2 handelt es sich um eine Weiterentwicklung des DVB-S-Standards. Derzeit befindet sich der vorgeschlagene DVB-S2-Standard in Begutachtung durch die ETSI (EN 302 307). Das neue Satelliten-Übertragungsverfahren DVB-S2 hat eine bis zu 30% höhere Effizienz gegenüber DVB-S. Weiter wurde die Codierung deutlich verbessert" [34].

### 2.3.7 DVB-C

DVB-C ist die kabelgebundene Variante des digitalen Fernsehens und soll hier nur kurz der Vollständigkeit wegen abgehandelt werden. Die Übertragung der Daten findet mittels den oftmals bereits vorhandenen Koaxial-Kabeln für die analoge Übertragung statt. Der Aufwand für die Umstellung auf digitale Übertragung hält sich in Grenzen. Für interaktive digitale Dienste muss das Kabelnetz rückkanalfähig sein. In der Schweiz ist die Aufrüstung auf rückkanalfähige digitale Kabelnetze weitestgehend abgeschlossen. Als Modulationsart kommt wie bei DVB-T QAM zum Einsatz. Der typische Durchsatz von rund 39 Mbit/s bei DVB-C ist dem von DVB-S ähnlich. Somit ist eine einfache Einspeisung von via DVB-S übertragenen Programmen ins Kabelnetz gewährleistet.

### 2.3.8 Vergleich der DVB Varianten

Tabelle 2.5 zeigt eine Aufstellung der spezifizierten Modulationsvarianten, der typischen verwendeten Bandbreite und die typischen jeweiligen Nutzdatenraten. Die Bandbreite

ist historisch bedingt, da in der etablierten analogen Frequenzbandbelegung 8 MHz für die Übertragung eines Fernsehkanals nötig sind. Die Übertragunskapazität hängt von der verwendeten Modulationsart, den externen Frequenzstörungen und von der Stärke der verwendeten Fehlerkorrektur ab.

Tabelle 2.5: Vergleich von Eigenschaften der DVB-Varianten.

| Merkmal | DVB-T | DVB-H | DVB-S | DVB-C |
|---|---|---|---|---|
| Modulation | 4, 16, 64 QAM | 4, 16, 64 QAM | 4 QAM | 16, 32, 64 QAM |
| Bandbreite | 8 | 8 | 36 | 8 |
| Datenrate | 20 | 14 | 38 | 38 |

### 2.3.9 IP Datacast

DVB-T, DVB-H, DVB-S sowie DAB Netzwerke können auch benutzt werden um IP Datenpakete an mobile oder fixe Terminals zu senden. Die nötigen Standards dazu existieren, z.B. DVB-DATA. Um dem Benutzer eine Interaktionsmöglichkeit zu geben sind Return Channels nötig. Bei DVB-S ist es möglich Antwortdaten direkt an Satelliten zu senden, DVB-T, DVB-H und DAB sind aber unidirektional spezifiziert. Als einzigen verlässlichen Return Channel bieten sich 2G oder 3G Netze an. Zurzeit werden in Europa zwei High Level Konzepte entwickelt für bidirektionales IP Datacasting in mobilen Umgebungen: Das IPDC Framework (Internet Protocol Datacasting Framework) und MBMS (Multimedia Broadcast Multicast Service). Die Gründe für Broadcasting und Multicasting sind:

- Bedürfnis nach Inhalt (z.B.Wetterdienste, Staumeldungen)
- Gleichzeitiges Empfangen von Inhalten
- Skalierbarkeit
- Höhere Datenraten als Unicast

#### Das IPDC Framework

IPDC ist kein Standard, sondern ein generisches, abstraktes Framework, das eine mögliche Organisation von mobilem bidirektionalem Datenverkehr beschreibt. Das IPDC Framework schreibt nicht vor welche physikalischen Trägermedien benutzt werden müssen, ist aber in erster Linie gedacht um via Broadcast zu senden und per 2G oder 3G Netzwerk zu antworten. IPDC gibt ein grobes technisches Organisationsschema vor, wie die ökonomische Wertschöpfungskette im Detail aussehen wird (Integrationsgrad, Agency Probleme), ist noch nicht vollständig geklärt und Gegenstand von Studien und kleinen Pilotprojekten in Skandinavien. IPDC via DVB-T ist nicht optimal, da für den DVB-T Empfang bei mobilen Geräten der Stromverbrauch im Verhältnis zur von der Batterie bereitgestellten Energie zu hoch ist. Auch haben mobile Geräte kleine Antennen, was zu erhöhten Fehlerraten bei der Übertragung von Daten führt. Deshalb sind ein Fehlerkorrekturalgorithmus sowie intensivere Übertragunssignale nötig. Für die Berücksichtigung der zusätzlichen

Einschränkungen wurde vom DVB-T Standard der DVB-H Standard abgeleitet, um einen auf die Bedürfnisse des mobilen Empfangs zugeschnittenen Standard verfügbar zu haben. DVB-H eignet sich gut als Broadcastkanal für mobiles IPDC. DVB-H ist wie DVB-T unidirektional konzipiert, für interaktive mobile Services kann als Return Channel eine 2G oder 3G Technologie verwendet werden. Das IPDC Framework angewendet auf DVB-T/H ergibt interessante technologische Möglichkeiten für mobile Breitband IP Services [11].

Die mögliche Bandbreite für den fixen IPDC Empfang ist 22 Mbps und rund 11 Mbps für den mobilen IPDC Empfang. Beim fixen Empfang von IPDC Services können die Daten in vorhandenen DVB-T gemultiplexten Transportströmen zusammen mit digitalen TV-Programmen übertragen werden. IPDC Transmissionen sind in der Regel als Broadcast vorgesehen, die Daten können von allen Clients innerhalb des Transmissionsradius empfangen werden. Um die Empfängergruppe einzuschränken wird im IPDC Framework Conditional Access (CA) sowie Digital Rights Management (DRM) berücksichtigt. Es ist möglich Daten zu verschlüsseln und Benutzerrechte den Daten, Streams oder Applikationen zuzuordnen [11].

Das IPDC Framework (Abbildung 2.7) wurde von IPDC-Forum vorgeschlagen. Zum Verständnis des IPDC Frameworks kann eine Gliederung der Layers (Core, Service Delivery, etc.) in 3 Einheiten, die "3 c's" vorgenommen werden [11]:

- Content: Beinhaltet die Erstellung und Lieferung von Services. Die involvierten Parteien sind Content Providers und Content Aggregatoren.
- Connection: Beinhaltet die Kern- und Zugangsnetzwerke der Netzwerkoperatoren.
- Consumption: End User Equipment (Hard- und Software).

Ein Layer kann aus verschiedenen physikalischen und logischen Komponenten bestehen. Die logischen Schnittstellen sind:

- Content Interface: Für die Lieferung des Content und dessen Beschreibung.
- Service Interface: Offeriert Funktionalitäten wie Streaming, Daten- und Applikationslieferung, Billing, DRM, Service Discovery und Service Request.
- Core Interface: Bietet QoS Garantien und IP Adressallozierungsfunktionalitäten.

In der Content Einheit befinden sich der Content- und der Service Layer. Im Content Layer befinden sich die Datenspeicherung, Cache Funktionen und ein Content Management System. Die für die Verrechnung von Diensten nötigen Daten werden hier erhoben. Digital Rights Management kann ebenfalls benutzt werden. Der Content Layer interagiert mit dem Service Deliverty Layer durch ein IP basiertes Netzwerk. Im Service Delivery Layer befinden sich Filter- und Aggregationsfunktionen für die Dienste. Session Management, Verfügbarkeit und Service Garantien sind im Service Delivery Layer definiert. Informationen für die Ankündigung von Diensten und für Return Channels werden hier generiert [11].

In der Connection Einheit sind Core, Broadcast- und der Return Channel Access Layers definiert. Der Core Layer ist ein multicast fähiges IP geroutetes Netzwerk, das die geforderte Netzwerkfunktionalität für IPDC zur Verfügung stellt. Billing Informationen werden gesammelt bezüglich der benutzten Bandbreite. Im Falle von Datenübertragung zu mobilen Endgeräten ist der Access Broadcast Layer IP basiert und unidirektional. Der Access Return Channel ist bidirektional und hat eine geringere Bandbreite als der Access Broadcast Layer. Die beiden Access Layers müssen auch Billing Informationen sammeln.

Die Consumption Einheit besteht aus dem Client Plattform und Application Layer und modelliert das Endgerät des Benutzers. Der Client Application Layer greift auf zur Verfügung gestellte Daten im Client Platform Layer via APIs zu. Im Client Plattform Layer befinden sich Zugangs-, Power Management-, Streaming-, Service Discovery-, DRM- und Zahlungsfunktionalitäten. Die Applikationen laufen im Client Application Layer. Eine logische Schnittstelle zum Content Layer garantiert das Funktionieren der Applikationen und Services über alle involvierten Layers hinweg.

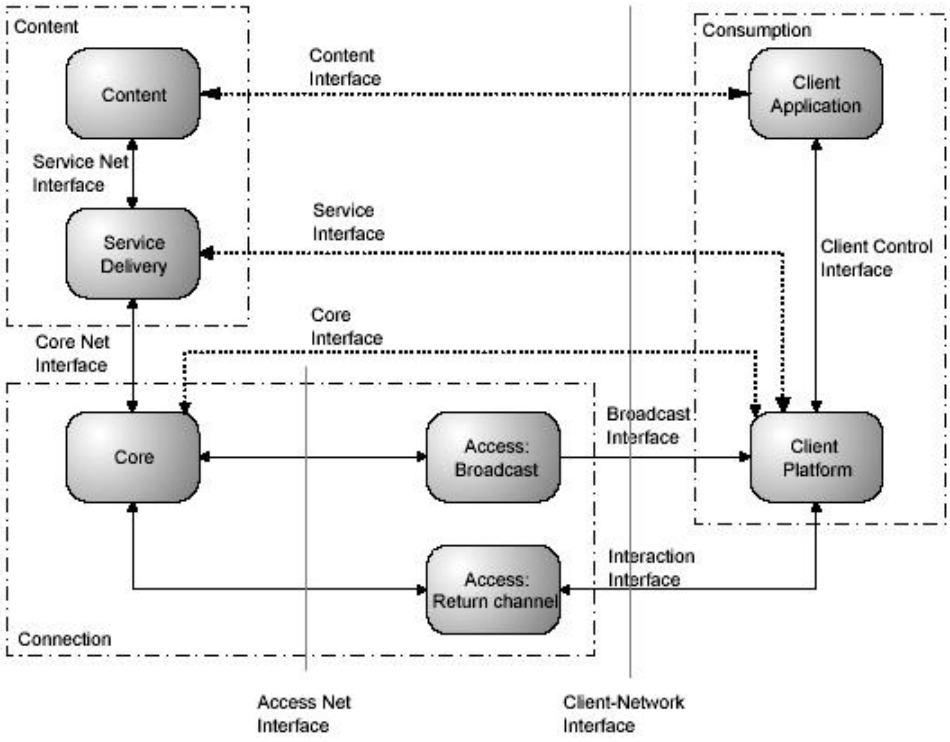


Abbildung 2.7: Das generische IPDC Framework. [11]

### MBMS

Multimedia Broadcast/Multicast Service (MBMS) ist vom Prinzip her ein auf dem IPDC Framework basierender Service, welcher mittels GSM oder UMTS angeboten werden kann. MBMS kann als Erweiterung von UMTS um eine Broadcastkomponente verstanden werden. Die technische Organisation von MBMS muss ebenfalls in der Lage sein die schon bei IPDC besprochenen Bedürfnisse (Billing, Content Aggregation, DRM, etc.) umzusetzen, allerdings Rücksicht nehmend auf die bereits vorhandenen 2G Netzwerkinfrastrukturen

respektive die schon spezifizierten 3G Netzwerke, da MBMS beim Senden und Antworten über ein Mobilfunknetzwerk kommunizieren soll. MBMS wird standardisiert durch 3GPP (Third Generation Partnership Project) und wird in Kürze in den UMTS Release 6 einfliessen. MBMS hat Potential, viele marktmächtige Telekomgesellschaften und Gerätehersteller haben sich am Standardisierungsprozess beteiligt. Es ist anzunehmen, dass in naher Zukunft MBMS Services erhältlich sein werden. MBMS ist gedacht für die Übertragung von kleinen Video- und Audioclips. Streaming ist beschränkt möglich, für grosse Datenvolumina (z.B. mobiles Fernsehen) eignet sich aber IPDC mit DVB-H besser. Schätzungen gehen davon aus, dass bis im Jahre 2010 rund 30% der Endgeräte MBMS fähig sind. Die Übertragung von Daten mit MBMS kann im Broadcast oder Multicast Modus geschehen. Die MBMS Architektur orientiert sich an der für den Mobilfunk vorhandenen Infrastruktur. Zentrales Element der Architektur ist das BM-SC (Broadcast Multicast Service Center). Im BM-SC wird MBMS koordiniert: Content Provider melden ihre Services an und die Lieferung des Contents wird organisiert [25]. Bild 2.8 zeigt den Broadcast von Content mit MBMS. Im Core Network fallen dank MBMS keine Point-to-Point Verbindungen mehr an. Für MBMS gibt es drei Service Szenarios: Streaming, "Download and Play" und Karussell. Anwendungen für MBMS sind z.B. Video und Audio Clips, lokalisierte Services (z.B. ein Touristeninformationskanal) und Service Portale.

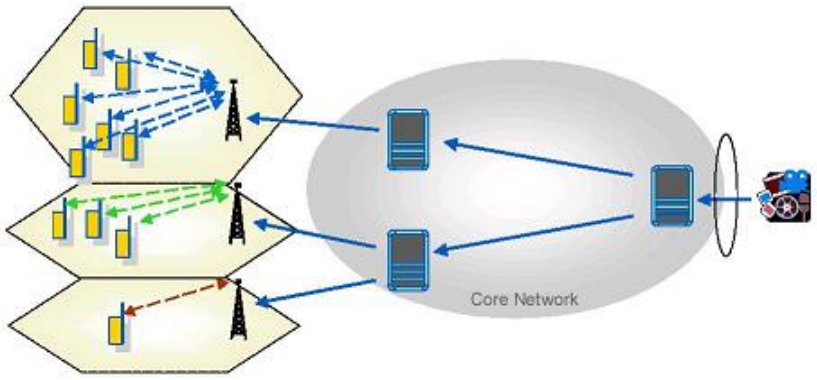


Abbildung 2.8: Die Übertragung von Content mit MBMS. [42]

### S-DMB

Satellite Digital Multimedia Broadcast (S-DMB) ist ein von der Firma Alcatel geführtes Projekt und soll hier als Beispiel dienen für die Möglichkeit Multicast Content direkt mittels Satellit an 3G kompatible Empfangsgeräte zu übertragen. Das S-DMB System sieht vor einen satellitengestützten multicast Layer über das terrestrische unicast 3G UMTS Netzwerk zu implementieren [35]. Die 3G fähigen Empfangsgeräte müssten in der Lage sein terrestrische und Satellitensignale zu empfangen. Im Prinzip können IP Daten mittels DVB-S gesendet werden, für den mobilen Empfang bei kleinen Endgeräten sind aber bei den Satelliten höhere Energieemisionen nötig. Das S-DMB System wird rund 1 Mbps Bandbreite pro Zone zur Verfügung stellen. Es ist vorgesehen Europa in 7 Zonen aufzuteilen. Der Vorteil ist, dass der Clearing Prozess der Frequenzregulation wesentlich einfacher sein wird als im Falle vom terrestrischen DVB. Die Empfangsgeräte müssten

viel Speicherplatz haben aus Gründen der Batterieeffizienz und der optimalen Bandbreitennutzung. Möglich wäre ein fortlaufendes Codieren von Videocontent mit 100 Kb/s, was einem kumulierten Datenvolumen von 430 MB/Tag entspricht. Ein S-DMB Sender könnte mit 144 KB/s im Karussell Modus senden und der empfangene Content wird bei den Endgeräten abgespeichert [32].

### 2.3.10 Gegenüberstellung der mobilen Datenübertragungstechnologien

Die drei oben vorgestellten Ansätze für mobiles Broadcasting mit Return Channel sind: IPDC basierend auf terrestrischem DVB, MBMS und S-DMB. Ihr gemeinsames Muster ist das Senden an den Empfänger im broadcast oder multicast Modus und ein logisch separater Return Channel, in der Praxis ein 2G oder 3G Mobilfunknetzwerk. Die Implementierung aller drei Ansätze verlangt komplexe technische Architekturen für das Management der angebotenen Services. Die Sendeinfrastruktur für IPDC via DVB ist vorhanden (Rundfunkantennen). Da MBMS ein Teil des UMTS Standards sein wird und als Treiber für wertschöpfende Dienste angesehen werden kann, sind bei MBMS die Sendeinfrastrukturkosten ebenfalls ökonomisch zu rechtfertigen. Im Falle von S-DMB müssten zusätzlich geostationäre Satelliten über Europa positioniert werden, was mit erheblichen Kosten verbunden wäre. Bild 2.9 stuft DVB-H und MBMS nach der Anzahl der erreichbaren Benutzer in Abhängigkeit der Servicepersonalisierung ein. Als Bezug werden die heutigen mobilen unicast Dienste angegeben. Tabelle 2.6 stellt die Leistungen der drei vorgestellten Broadcasttechnologien gegenüber [36]. "+" bedeutet "geringe Leistung" und "+++" bedeutet "hohe Leistung".

Tabelle 2.6: Leistungsvergleich MBMS, S-DMB und IPDC.

| Merkmal | MBMS | S-DMB | IPDC |
|---|---|---|---|
| Bandweite | ++ | + | +++ |
| Batterieeffizienz | +++ | ++ | ++ |
| Kernnetzkomplexität | + | +++ | ++ |
| Live Content (dynamische Bandweite) | ++ | + | +++ |
| Geogr. Reichweite | + | +++ | ++ |
| Lokale Granularität | +++ | + | ++ |
| Integration mit Unicast-Diensten | +++ | + | ++ |
| IP Meta-Data and Mikro Marketing Kompatibilität | +++ | + | ++ |

### 2.3.11 Konvergenz der mobilen Datenübertragung

Die mobilen Datenbroadcasttechnologien entstehen aufgrund der Annahme einer grossen Nachfrage nach mobiler Datenübertragung der Benutzer und der Tatsache, dass diese Nachfrage aufgrund beschränkter zur Verfügung stehender Frequenzbänder nicht mehr mit traditionellen unicast Systemen zu befriedigen ist. Aus Konsumentensicht ist die Zukunft des Broadcasting "Interactive Multemdia Delivery" - zu erschwinglichen Preisen ohne sich

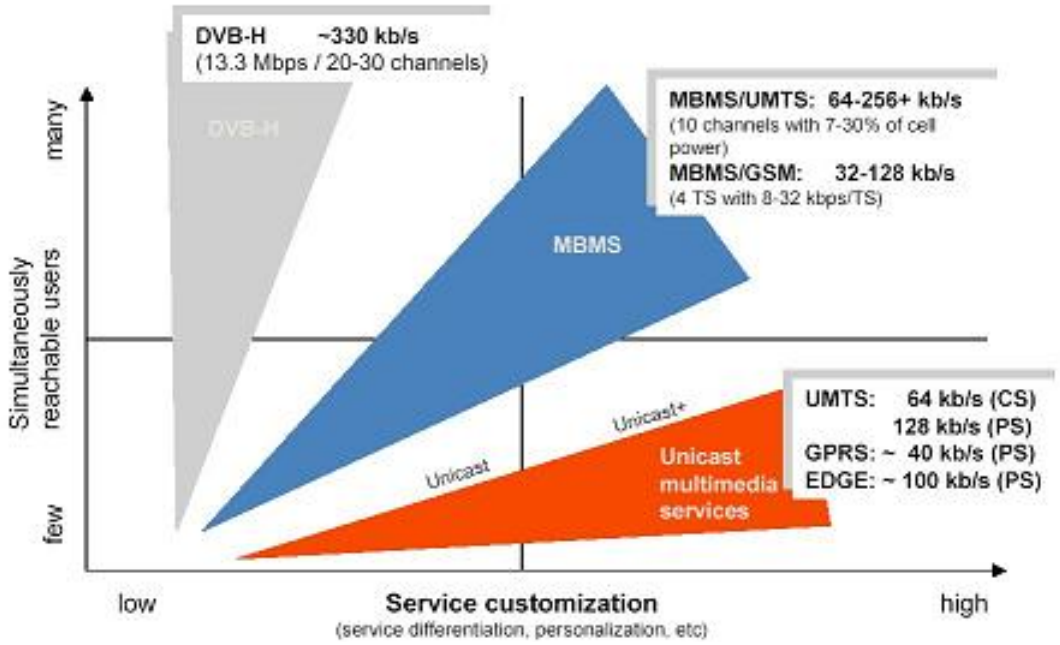


Abbildung 2.9: Performance Vergleich der technischen Optionen. [42]

sorgen machen zu müssen wegen Heterogenität der Standards und Technik. Um zu dieser idealen Kommunikationslandschaft zu kommen, müssten die in diesem Paper erläuterten Broadcasttechnologien mit den 2G und 3G Netzen sinnvoll kombiniert werden. Es müsste eine Konvergenz der Services und der Netzwerke stattfinden. Genauer genommen müssten Services (Unterhaltungsindustrie), Stysteme und Netzwerke der Broadcast-, Telekom- und Computerindustrie näher zusammenrücken. Konvergenz heisst nicht, dass eine Industrie aus einer Hand alle Bedürfnisse im Zusammenhang mit Broadcast befriedigen soll, sondern Kooperationen unter den Beteiligten eingegangen werden müssten um letztlich ein breites Spektrum an Services und Terminals den Konsumenten anbieten zu können. Den Kooperationen stehen ökonomische sowie kulturelle Probleme im Wege. Beispielweise werden Telekomgesellschaften nur schwer zu überzeugen sein Downstream Traffic via DVB-H im IPDC Framework zuzulassen, weil sie viel Geld für UMTS Lizenzen bezahlt haben und an Traffic respektive Umsatz über ihre eigenen Netze interessiert sind (MBMS). Auch müssten sich Unterhaltungs- und Telekommunikationsindustrie näher kommen im Interesse der Konsumenten. Damit diese beiden Branchen zusammenfinden, sind aber noch beträchtliche kulturelle Hürden zu nehmen. Aus heutiger Sicht kann die ideale interaktive mobile Multimedia Services Broadcastingzukunft so geschildert werden [3]:

- Downstream Übertragung mittels digitalen Broadcastnetzwerken.
- Empfängergeräte mit Speicher- und Verarbeitungsfunktionen für Interaktivität und Zugang zu Programmen nach Wunsch des Benutzers.
- Upstream Übertragung mittels Mobilfunknetzwerken.

## 2.4 Schlusswort

Die vorliegende Arbeit hat in den europäischen digitalen Rundfunk eingeführt. Ziel war es, die Standards zu erklären und die Anwendungen aufzuzeigen, einen Überblick zu der tatsächlichen Verbreitung zu geben und die verwendeten Technologien offen zu legen.

DAB ist ein europäischer Standard basierend auf MPEG-2 zur Übertragung von Radio in digitaler Form. DAB verspricht Klangqualität vergleichbar mit einer Audio CD und soll bis in Jahr 2010 das analoge Radio ablösen. Zusatzinformationen wie z.B. der Titel des gerade gespielten Liedes können ebenfalls mittels des DAB Standards übermittelt werden. DRM ist ebenfalls ein digitaler auf MPEG-4 basierender Radio Rundfunkstandard. DRM ist wie DAB ein offener Standard und wurde ins Leben gerufen von einem Konsortium, dass sich auf das weltweite Senden (mittels Langwelle) von Radiosendungen konzentriert. DAB und DRM liegen nicht in direkter Konkurrenz, sie ergänzen sich.

Hinter dem Kürzel DVB verbergen sich rund 100 Standards, die alle Aspekte des digitalen Sendens von Informationen spezifizieren. Die massgebenden Standards sind DVB-Terrestrial, DVB-Handheld, DVB-Satellite und DVB-Cable. Sie definieren die Übertragung von Datenströmen, in erster Linie Audio und Video für digitalen Rundfunk über verschiedene physische Trägermedien bzw. Distanzen.

Die digitalen Rundfunkstandards haben klare Vorteile und werden über kurz oder lang eingeführt werden. Die von verschieden europäischen Ländern gesetzten ehrgeizigen Zieldaten für die definitive Einführung der digitalen Rundfunkstandards werden voraussichtlich nicht eingehalten werden können, da die Konsumenten zurzeit nicht genügend Anreize haben auf die neuen Technologien umzusteigen. DVB war ursprünglich gedacht als Standard für digitales Fernsehen, wird jedoch immer wichtiger für IP Datenübermittlung bei mobilen Systemen, weil potentiell eine hohe Bandbreite günstig zur Verfügung steht. Der jüngste Standard, DVB-H wurde auf die Bedürfnisse von mobilen Benutzern zugeschnitten: Hohe mögliche Datenraten bei Empfängern, die sich bewegen unter Berücksichtigung der beschränkten Batterielaufzeit.

Mit der rasanten Entwicklung der Leistungfähigkeit von mobilen Endgeräten ist mobiles Breitband in Griffweite. Das generische IPDC Framework schlägt eine Organisation der zukünftigen mobilen bidirektionalen Breitband Services vor. IPDC wird vor allem im Zusammenhang mit DVB-H als Downstream Channel erwähnt. MBMS ist die Broadcasterweiterung der Telekommunikationsindustrie, der gesamte Datenverkehr läuft über Mobilfunknetzwerke. S-DMB ist ein Projekt, welches die Broadcastkomponente für Mobilfunknetzwerke via Satellit zur Verfügung stellen möchte.

# Literaturverzeichnis


[1] Welcome to world DAB, http://www.worlddab.org/eureka.aspx, März 2005.

[2] Digital Audio Broadcasting - Wikipedia, http://en.wikipedia.org/wiki/Digital_Audio_Broadcast, März 2005.

[3] Jan Doeven: A road map to broadcast technology, http://www.ebu.ch/trev_294-doeven.pdf, März 2005.

[4] Digital Radio Mondiale - Wikipedia, http://en.wikipedia.org/wiki/Digital_Radio_Mondiale, April 2005.

[5] Digital Radio Mondiale, http://www.drm.org/system/globtechnical.htm, März 2005.

[6] Peter Röbke-Doerr: Digitales Radio, http://www.heise.de/ct/03/16/028/default.shtml., April 2005.

[7] Hans Weber: Broadcast Magazine, http://www.drm-national.de/DRM_vor_Durchbruch.pdf, März 2005.

[8] China News Release, http://www.drm-national.de/China_Newsrelease.pdf, April 2005.

[9] DVB - Wikipedia, http://en.wikipedia.org/wiki/DVB, März 2005.

[10] DVB - Digital Video Broadcasting - White Papers, http://www.dvb.org/index.php?id=20, März 2005.

[11] Sonera MediaLab, http://www.ipdc-forum.org/resources/documents/IPDCContentServicesWP.pdf, März 2005.

[12] Digitalfernsehen.de, http://www.digitalfernsehen.de, April 2005.

[13] Bakom - digitales terrestrisches Fernsehen auch für die Schweiz, http://www.bakom.ch/de/radio_tv/dvb/dvb_t/index.html, April 2005.

[14] Einführung von DVB-T, http://www.srgssrideesuisse.ch/de/tv/reception/de_dvb.html, März 2005.

[15] Hans Strassmann: DVB-T in der Schweiz - Technische Aspekte, http://www.srgssrideesuisse.ch/de/home/290103/strassmann.ppt, März 2005.

[16] Vorlesung Medientechnik, http://tech-www.informatik.uni-hamburg.de/lehre/ss2004/vorlesungen/medientechnik/09-dvbt.pdf, April 2005.

[17] Rainer Keller: DVB-T - Perspektiven der SRG SSR, http://www.srgssrideesuisse.ch/de/home/290103/keller.ppt, April 2005.

[18] Dr. Volker Zota: Auf DVD gesendet, DVB-T-Aufnahmen verlustfrei bearbeiten und auf DVD archivieren, S. 122., c't 11/04.

[19] Frequenzband - Wikipedia, http://de.wikipedia.org/wiki/Frequenzband, April 2005.

[20] Banbreite - Wikipedia, http://de.wikipedia.org/wiki/Bandbreite, April 2005.

[21] MPEG-2 Transmission, http://www.erg.abdn.ac.uk/research/future-net/digital-video/mpeg2-trans.html, April 2005.

[22] Torsten Jaekel: Einführung in DVB, http://home.t-online.de/home/Torsten.Jaekel/dvb.htm, März 2005.

[23] DVB - Digital Video Broadcasting, http://www.dvb.org/index.php?id=27, Mai 2005.

[24] Alexander Braun Markus Hofbauer: Semesterarbeit über digitales Fernsehnen, http://people.ee.ethz.ch/ãmbraun/sa1/kapitel2.html, Mai 2005.

[25] Mobile Broadcast/Multicast Service (MBMS) White Paper, http://www.medialab.sonera.fi/workspace/MBMSWhitePaper.pdf, Mai 2005.

[26] Das überall Fernsehen - Technik, http://www.ueberall-tv.de/3content/tech/technik.htm, Mai 2005.

[27] Phasenmodulation - Wikipedia, http://de.wikipedia.org/wiki/Phasenmodulation, Mai 2005.

[28] Quadraturamplitudenmodulation - Wikipedia, http://de.wikipedia.org/wiki/QAM, Mai 2005.

[29] Kabelkommunikation: Digitales Fernsehen, http://www.kefk.net/Research/Kabelnetze/digitales_fernsehen.html#DF1 Mai 2005.

[30] Discrete Cosine Transform, http://en.wikipedia.org/wiki/DCT, Mai 2005.

[31] COFDM, http://de.wikipedia.org/wiki/COFDM, Mai 2005.

[32] Moby Dick Summit, http://www-ks.rus.uni-stuttgart.de/Events/031106_MobyDick-Summit/03-MobyDick-Summit-031106-MoDiS-Workshop.pdf, Mai 2005.

[33] Internet Magazin, http://www.internet-magazin.de/praxis/cm/page/page.php?table=pg&id=1319, Mai 2005.

[34] DVB-S, http://de.wikipedia.org/wiki/DVB-S, Mai 2005.

[35] ESA Telecommunications, http://telecom.esa.int/telecom/www/object/index.cfm?fobjectid=11985, Mai 2005.

[36] J. Oliver, D. Tymen: Scaleable Mobile Media Delivery: DVB-T and DVB-S to become the Bedfellows of UMTS, http://www.broadcastpapers.com/sigdis/ibc2003UDcastUMTS.pdf, Mai 2005.

[37] WIPO Copyright Treaty, http://www.wipo.int/treaties/en/ip/wct/trtdocs_wo033.html, Mai 2005.

[38] DVB-T, http://www.dvb-t4me.de/dvbt_wie_funktioniert_das.php, Mai 2005.

[39] DVB-H, http://www.dvb.org/documents/white-papers/wp07.DVB-H.final.pdf, Mai 2005.

[40] Digital Audio Broadcasting, http://www.srg.ch/de/radio/distribution/dab.html, Mai 2005.

[41] MPEG-2 Transmission, http://www.erg.abdn.ac.uk/research/future-net/digital-video/mpeg2-trans.html, Juni 2005.

[42] Mobile Broadband, http://www.ericsson.com/de/presse/background/ Ericsson_ims_hsdpa_mbms_01_2005.pdf, Juni 2005.

# Kapitel 3

# Technology and Use of RFID

*Manuel Ziltener und Elias Diem*

*Diese Seminararbeit umfasst die RFID (Radio Frequency Identification) Technologie sowie mögliche Anwendungsgebiete. Nach der Klärung technischer Grundlagen werden Anwendungen vorgestellt, die auf der RFID Technologie beruhen. Diese Anwendungen decken ein breites Spektrum ab, sind aber nicht als vollständige Aufzählung gedacht. Aufgrund der bereits im Einsatz stehender Anwendungen und solchen, die erst noch breitere Verwendung finden müssen, soll kritisch über die Technologie diskutiert werden können. Es werden dazu Chancen und Risiken dieser Technologie aufgezeigt.*

## Inhaltsverzeichnis

## 3.1 Einleitung

Die Identifizierung von Objekten ist schon immer ein wichtiges Thema gewesen. Zum Beispiel ist der Barcode aus allen Lebensmittelläden bekannt [1]. Mit der RFID (Radio Frequency Identification) Technologie steht erstmals eine Technologie zur Verfügung, welche es ermöglicht, jedem Objekt auf der ganzen Welt eine eindeutige Nummer zur Verfügung zu stellen und diese Objekte auch global zu verwalten und zu lokalisieren. Ferner gibt es Unterschide zu herkömmlichen Identifizierungssystemen, die vor allem die Speicherkapazität, die Robustheit und die Lesbarkeit betreffen.

Die RFID Technologie wird von verschiedenen Unternehmen als grosse Chance wahrgenommen, um Kosten einzusparen. Die internen Prozesse sollen weiter automatisiert und Fehlerquellen minimiert werden. Vorangetrieben wird diese Technologie vor allem durch grosse Handels- und Logistikkonzerne, die ihre Produkte über die gesamte Wertschöpfungskette verfolgen wollen.

Auf Konsumentenseite wird die Technologie eher noch mit Misstrauen angeschaut. Dieses Misstrauen bezieht sich vor allem auf den möglichen Verlust der Privatsphäre. Dieses Verhaltensmuster wird immer bei der Einführung oder dem Wechsel neuer Technologien beobachtet.

In einem ersten Teil wird auf die Technologie eingegangen, die RFID zugrunde liegt. In einem zweiten Teil werden verschiedene Anwendungsbeispiele aufgezeigt. Dabei soll das technologische Wissen, welches diesen Beispielen vorangestellt wurde, helfen, Vor- und Nachteile gegenüber anderen Technologien abzuschätzen.

## 3.2 Die RFID Technologie

Der ersten Teil beginnt mit einer allgemeinen Einführung in die RFID Technologie. Danach wird auf die geschichtliche Entwicklung der RFID Technologie eingegangen, um dann das Auto-ID Center vorzustellen, dass eine wichtige Rolle bei der ganzen RFID Technologie spielt. Die Funktionsweise von RFID Tags und die Standard, die dazu existieren, werden danach behandelt. Zum Schluss wird der EPC (Electronic Product Code) vorgestellt.

### 3.2.1 Einführung in RFID

Am Beginn des Einführungsteils soll eine Definition von RFID stehen:

RFID (Radio frequency identification) ist eine Technologie, die ein System beschreibt, dass die Identität eines Objektes (das können auch Personen oder Tiere sein) kabellos mittels Radiowellen übermittelt. Zur Identifizierung wird eine eindeutige Nummer verwendet [2].

Generell kann gesagt werden, dass Identifizierungssysteme Barcodes, optische Zeichenleser und biometrische Technologien (zum Beispiel Irisscaner) beinhalten. Bei Barcodesystemen werden meistens Personen benötigt, welche die Daten einlesen. Das funktioniert bei RFID

anders, weil Scaner die Daten auf Distanz ohne menschliches Zutun einlesen können. Es braucht auch keinen direkten Sichtkontakt. Die Daten werden dabei auf so genannten RFID Tags gespeichert.
Ein solcher RFID Tag besteht aus einem Microchip, der an einer Radioantenne angeschlossen ist. Der Microchip kann verschiedene Volumen von Daten speichern (1 bit bis 1 MByte [3]). Es gibt die Möglichkeit die Daten nur auszulesen oder aber sie zu ändern. Um an diese Daten zu gelangen, braucht es ein Lesegerät, das typischerweise mit einer oder mehreren Antennen ausgestattet ist. Über diese Antennen werden Radiosignale ausgesendet, die vom RFID Tag empfangen werden und gewisse Signale zurückgesendet werden. Die Lesegeräte leiten diese Daten an ein Computersystem weiter.

Hier kommt die Frage auf, was denn nun ein RFID System von einem Barcode System unterscheidet.

#### Unterschiede zwischen RFID und Barcode

Im nachfolgenden Abschnitt wird versucht die RFID Technologie von den weit verbreiteten Barcodes abzugrenzen und so Vorteile und Nachteile beider Technologien aufzuzeigen. Die RFID Technologie wurde unter anderem entwickelt, um den Barcode abzulösen. Vier Hauptunterschiede werden aufgelistet.

1. Der Barcode besitzt heute für jede Produktgruppe, zum Beispiel Coca Cola eine EAN-Nummer [1]. Wie weiter unten noch beschrieben wird, besitzt RFID jedoch für jedes physisch produzierte Gut auf dem Planet eine eigene Nummer, mit welcher diese Güter eindeutig identifizierbar, katalogisierbar und verfolgbar sind. RFID ist somit optimal bei der Identifikation von Objekten.

2. Ein zweiter markanter Unterschied ist, das RFID anders als Barcode, aus Entfernung gescannt werden kann, also ohne direkten Sichtkontakt. Dies kann anders als bei den Barcodelesegeräten auch verdeckt stattfinden, ohne dass sich eine betroffene Person dagegen wehren kann.

3. Der dritte Unterschied ist die vorhandene Speicherkapazität. RFID Tags besitzen eine wesentlich höhere Speicherkapazität als der etablierte Barcode.

4. Weiter ist die Erfassung von Barcode nur bei langsam bewegten oder ruhenden Objekten möglich, wobei der Barcode auch nicht verdeckt sein darf. Bei RFID spielt die Geschwindigkeit des Objektes weniger eine Rolle, da die Datenübertragung eine sehr hohe Geschwindigkeit besitzt und der Tag kann verdeckt sein.

### 3.2.2 Geschichte von RFID

Um den geschichtlichen Aspekt der RFID Technologie aufzuzeigen, werden in diesem Teil technische Aspekte angesprochen. Es wird jedoch auch auf RFID Anwendungen eingegangen, die zum Teil im Anwendungsteil der Arbeit weiter vertieft werden. Die wesentlichen Informationen sind [5] entnommen.

Die Geschichte von RFID beginnt schon im zweiten Weltkrieg. Die Deutschen, Japaner, Amerikaner und Briten benutzten alle Radar, das 1935 von Robert Alexander Watson-Watt entdeckt worden war. Es wurde dazu verwendet, Flugzeuge zu orten auch wenn sich noch sehr weit weg waren. Das Problem dabei war, dass nicht zwischen den eigenen und den feindlichen Flugzeugen unterschieden werden konnte. Die Deutschen fanden jedoch heraus, dass das reflektierte Signal anders war, wenn sie mit dem Flugzeug rollten. Diese Art der Datenübertragung kann als erstes passives RFID System angesehen werden. Unter Watson-Watt arbeiteten die Briten an einem System zur Freund/Feinderkennung. Sie statteten die Flugzeuge mit Transmittern aus. Wenn von einer Radarstation am Boden ein Signal empfangen wurde, wurde das Signal entweder reflektiert (entspricht einem passiven System) oder es wurde ein eigenes Signal zurück gesendet (entspricht einem aktiven System).

Die Entwicklung von RFID Systemen ging weiter und es wurde vermehrt an Diebstahlsicherungssystemen und Produkteüberwachung gearbeitet. Bekannt sind heute sicher die Diebstahlsicherungen in den Läden. Dabei wird ein RFID Tag mit einem Bit verwendet. Ist er eingeschaltet, d.h. das Bit ist gesetzt, geht ein Alarm beim Verlassen des Ladens los. Im Jahr 1973 wurden die ersten RFID Patente angemeldet.
Zu dieser Zeit arbeitete auch die US Regierung an RFID Systemen, vor allem um nukleares Material zu verfolgen. Im Los Alamos National Laboratory wurde deswegen geforscht. Es wurde ein System entwickelt, dass einen vorbeifahrenden Lastwagen anhand dessen Transponder erkennen konnte. Diese Anwendung wurde dann Mitte der achtziger Jahre in automatischen Zahlsystemen weiterentwickelt, die auf Autobahnen, Brücken oder Tunneln eingesetzt wurden.
Des weiteren entwickelte das Los Alamos National Laboratory ein System zur Identifikation von Kühen. Dabei wurde den Kühen ein passiver RFID Tag unter die Haut injiziert. Dieses System benutzte Radiowellen im 125 kHz Bereich. Zuerst wurde dieser 125 kHz Bereich noch von anderen kommerziellen Produkten verwendet, bevor dann auf Frequenzen um 13.56 MHz (high frequency) umgestiegen wurde. Dies weil dieses Band in den meisten Ländern nicht reguliert war. Andere Anwendungen in diesem Band sind Zutrittskontrollen, Zahlungssysteme oder Diebstahlsicherungen in Autos.
Anfangs der neunziger Jahre wurde von IBM im UHF (ultra high frequency) Band RFID entwickelt, was grössere Distanzen vom Tag zum Leser und schnellerer Datentransfer mit sich brachte. Diese Systeme konnten sich aber noch nicht durchsetzen, weil es noch zu teuer war und keine internationalen Standards existierten.

UHF RFID bekam einen Schub im Jahr 1999 als das so genannte Auto-ID Center gegründet wurde. Bis 2003 traten über 100 grössere Unternehmen dem Auto-ID Center bei, darunter auch RFID Hersteller. Es wurden Forschungslabors in Australien, England, der Schweiz, Japan und China eröffnet.

### 3.2.3 Das Auto-ID Center

Wie im letzten Abschnitt angesprochen, spielt das Auto-ID Center in der Entwicklung der RFID Technologie eine wichtige Rolle [5]. Es wurde am Massachusetts Institute of Technology gegründet. Die Gründer waren das Uniform Code Council (UCC), EAN International,

Gillette und Procter & Gamble. Das Center verfolgt primär zwei Ziele. Das erste Ziel ist mit der Privatwirtschaft zusammen zu arbeiten und einen RFID Tag zu entwickeln, der extrem billig ist. Der Zielpreis war fünf US-Cent. Der tiefe Preis ist wichtig, damit ein RFID Tag nach einmaligem Gebrauch weggeworfen werden kann. Die Übertragung der Daten musste im UHF Bereich erfolgen, weil nur so aus den gewünschten Entfernungen die RFID Tags ausgelesen werden konnten. Der fünf US-Cent RFID Tag ist allerdings noch nicht Realität geworden, gegenwärtig kostet er zwischen 20 und 40 US-Cent [9].
Es kommen verschiedene Strategien in Frage, wie nachfolgend aufgezeigt, wenn man den Preis auf die fünf US-Cent bringen will [8].

- Herstellungskosten durch grössere Stückzahlen senken
- Antenne verkleinern um dadurch den Kupferanteil senken
- Günstigeren aber trotzdem ebenso robusten Kunststoff für den ganzen Tag einsetzen
- Microchip vereinfachen und billigeren Herstellungsprozess verwenden

Das Auto-ID Center beschäftigte sich aber nicht nur mit der Entwicklung eines billigen RFID Tags, sondern baute auch EPC (Electronic Product Code, siehe weiter unten) auf, ein Nummeringungsschema, welches es möglich macht, eine Seriennummer auf jedes hergestellte Produkt zu platzieren. Ferner wurde ein Kommunikationsprotokoll entwickelt, mit welchem RFID Tag und Lesegerät kommunizieren können und ein sicherer Weg, um die Daten über das Internet abzuspeichern.
Der EPC stellte den entscheidenden Unterschied zu den bis dahin verwendeten RFID Systemen dar. Es wird eine eindeutige Seriennummer auf jedem Produkt gespeichert. Diese Nummer hat keinen grossen Platzbedarf. Mittels der Nummer kann nun in einer Datenbank über das Internet alle Information über das Produkt abgefragt werden, ohne dass diese Information auch auf dem RFID Tag gespeichert werden müsste. Bei der Herstellung von RFID Tags ist das der entscheidende Kostenvorteil. Je weniger Information auf ihm gespeichert werden muss, desto billiger ist er herzustellen.

Zu dem Center stiessen im Verlaufe der Zeit immer mehr amerikanische Unternehmen wie zum Beispiel Kimberly-Clark, Metro, Target, Tesco, Unilever und Wal-Mart. Im Jahr 2003 spaltete sich das Auto-ID Center in zwei Teile auf. Es gab neu das Auto-ID Labs, das weiterhin am MIT forschte und das vor allem im Bereich von EPC. Die EPC Technologie wurde an das Uniform Code Council lizenziert, welches EPCglobal in Zusammenarbeit mit EPC International gründete.

### 3.2.4 Funktionsweise von RFID

Die allermeisten RFID Systeme benutzen einen Microchip aus Silikon, um darauf eine eindeutige Seriennummer zu speichern. Meistens werden noch weitere Daten darauf gespeichert. Es gibt zwei Hauptunterscheidungsmerkmale der RFID Tags: aktiv und passiv. Aktive RFID Tags haben einen eigenen Transmitter und eine Energiequelle, meist in Form einer Batterie. Eine alternative Energiequelle wäre zum Beispiel eine Solarzelle. Aktive

RFID Tags senden ein Signal aus, um die gespeicherte Information zu übertragen. Ein typischer aktiver RFID Tag ist in Abbildung 3.1 dargestellt [6]. Dieser aktive RFID Tag ist in der Lage, Gammastrahlung zu messen.

Abbildung 3.1: Ein aktiver RFID Tag

Passive RFID Tags haben keinen Transmitter, sie reflektieren lediglich die empfangene Energie, die vom Lesegerät kommt. Ein typischer passiver RFID Tag ist in Abbildung 3.2 dargestellt [7]. In der Mitte befindet sich der Microchip aus Silikon (schwarz). An ihn angeschlossen und aufgewickelt ist die Antenne zu sehen.

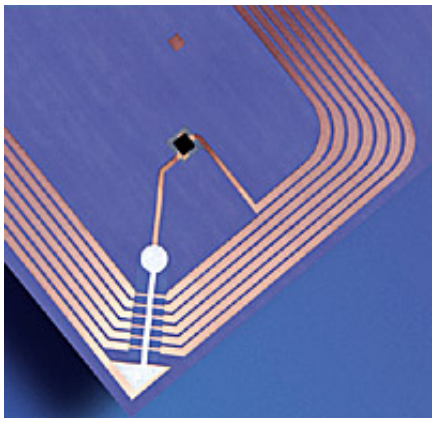

Abbildung 3.2: Ein passiver RFID Tag

### Aktive RFID Systeme

Aktive RFID Tags werden auf grösseren Einheiten gebraucht, wie zum Beispiel Verladekontainern oder Bahnwagons. Sie arbeiten normalerweise im 455 MHz, 2.45 GHz oder 5.8 GHz Band und haben eine Reichweite von 20 bis 100 Metern [9]. Aktive RFID Systeme werden nochmals in zwei Gruppen unterteilen: Transponder und Beacons (auf Deutsch: Ortungsgerät).

Transponder arbeiten normalerweise nicht, sondern werden aufgeweckt, sobald sie ein Signal eines Lesegerätes empfangen. Ein Anwendungsgebiet dafür sind zum Beispiel die

Abrechnungssysteme auf den Autobahnen. Nachdem der RFID Tag das Signal des Lesegerätes empfängt, wird er aktiviert und sendet seine eindeutige Identifikationsnummer an das Lesegerät.

Anders arbeiten die Beacons. Sie senden permanent und werden daher zum Beispiel in Echtzeitlokalisierungssystemen (RTLS: Real-Time Locating System) gebraucht. Bei solchen Systemen muss die genaue räumliche Lage einer Einheit zu jedem Zeitpunkt bekannt sein. Ein Anwendungsgebiet sind zum Beispiel grosse Lager wie auf Abbildung 3.3 gezeigt. Diese RFID Tags senden daher in einem festgelegten Zeitintervall Signale ab. Wenn dann dieses Signal von mindestens drei Lesegeräten registriert wird, kann die genaue Position der Einheit berechnet werden.
Ein aktiver RFID Tag kostet zwischen 10 bis 50 US-Dollar, je nachdem wie viel Information er speichern kann, wie lange die Batterie hält und wie robust er gebaut ist [9].

Abbildung 3.3: Ein grosses Containerlager

**Passive RFID Systeme**

Passive RFID Tags besitzen keine Batterie und keinen Transmitter. Sie sind billiger als aktive RFID Tags und benötigen keinen Unterhalt. Die Auslesedistanz ist kleiner, sie geht von ein paar Zentimetern bis zu 10 Meter [9].
Passive RFID Tags können im LF (low frequency), HF (high frequency) und UHF (ultra high frequency) Band betrieben werden. Im LF Band verwenden sie Frequenzen von 124 kHz bis 135 kHz. Im HF Band sind es 13.56 MHz und im UHF Band liegen sie zwischen 860 MHz bis 960 MHz.
Da Radiowellen sich bei unterschiedlichen Frequenzen verschieden verhalten, muss je nach Anwendung entschieden werden, in welchem Band die RFID Tags betrieben werden.
Generell kann gesagt werden, dass kurze Wellen sich wie die bekannten Radiowellen verhalten, also leicht Wände durchdringen, Metall aber nicht. Kurze Wellen dagegen ähneln eher dem Licht und können Materialien nicht so gut durchdringen. Ferner werden sie an vielen Orten reflektiert.
Die Herausforderung der RFID Tag Hersteller besteht also darin, die Tags auch auf metallischen Oberflächen, in der Umgebung von Flüssigkeiten oder inmitten von vielen anderen Einheiten zu lesen. Die Lesedistanz hängt dabei stark mit dem verwendeten Band zusammen.
Im UHF Band können dabei die Distanzen bis 10 Metern erreicht werden. Im LF Band dagegen kann nur bis 30 Zentimetern ausgelesen werden.

Die Lesedistanz wird aber auch noch von einem anderen wichtigen Faktor beeinflusst, nämlich der Art und Weise wie die passiven RFID Tags Daten übermitteln. Es wird zwischen zwei Arten unterschieden. Die LF und HF RFID Tags verwenden das so genannte induktive Koppelung, während UHF RFID Tags die Sendekoppelung verwenden.

Die induktive Koppelung funktioniert folgendermassen: je eine Drahtspule im RFID Tag und im Lesegerät bilden zusammen ein elektromagnetisches Feld. Der RFID Tag bezieht Energie von diesem Feld, braucht sie, um seinen Microchip zu betreiben und ändert dann die Spannung an der Antenne. Die Antenne im Lesegerät registriert diese Änderungen des Magnetfelds und übersetzt sie in Null und Eins. Weil die zwei Drahtspulen ein elektromagnetisches Feld aufbauen, darf die Distanz zwischen ihnen nicht zu gross sein.

Bei der Sendekoppelung wird kein elektromagnetisches Feld aufgebaut. Das Lesegerät sendet Radiowellen aus. Der RFID Tag braucht diese Wellen, um Energie zu beziehen. Diese Energie wird gebraucht, um die Last an der Antenne zu ändern und damit ein verändertes Signal zurück zu senden. Diesen Vorgang wird auch als backscatter bezeichnet.

UHF RFID Tags können mittels drei verschiedenen Modulationsarten ihre Daten übertragen. Diese Arten sind: die Amplitudenmodulation (die Stärke des Signals wird verändert), die Frequenzmodulation (die Frequenz der Wellen wird geändert) und die Phasenmodulation (die Phase wird verschoben).

Die Kosten für einen passiven RFID Tag liegen gegenwärtig bei 20 bis 40 US-Cent [9].

Einen Überblick der einzelnen technischen Aspekte des RFID Tags und drei weiteren Technologien liefert die Tabelle 3.1.

### Leistungsuntersuchungen

Weil LF und HF RFID Tags die induktive Koppelung benutzen, ist der Lesebereich kleiner, was ihn einfacher zu kontrollieren macht. Bei den UHF RFID Tags ist der Lesebereich viel grösser, weil weitere Distanzen überwunden werden müssen. Die Radiowellen können an vielen Orten reflektiert werden und so RFID Tags erreichen, die eigentlich gar nicht erreicht werden wollen. Somit werden auch mal RFID Tags gelesen, die gar nicht gelesen werden wollen.
LF und HF RFID Tags arbeiten besser in der Umgebung von Metall und Wasser. UHF Wellen werden zum Beispiel von Wasser absorbiert. Wenn Getränke mit RFID Tags versehen werden, muss dies berücksichtigt werden. Es kann zum Beispiel ein Zwischenraum zwischen der Flüssigkeit und dem RFID Tag geschaffen werden. Der Zwischenraum ist auch zwischen dem RFID Tag und Metall nötig. Auch durch Interferenz wird die Performance beeinflusst. Darunter wird so genanntes Rauschen verstanden, das von vielen anderen Geräten abgegeben werden kann (zum Beispiel ein laufender Motor oder noch stärker ein Starten des Motors) und es schwierig macht, das interessierende Signal zu erkennen.

Tabelle 3.1: Vergleich der Technologien

| Parameter / System | Barcode | OCR (Optical Character Recognition) | Chipkarte | RFID |
|---|---|---|---|---|
| Typische Datenmenge (Byte) | 1- 100 | 1- 100 | 18 - 64k | 1 - 1M |
| Datendichte | Gering | Gering | Sehr hoch | Sehr hoch |
| Maschinenlesbarkeit | Gut | Gut | Gut | Gut |
| Lesbarkeit durch Personen | Bedingt | Leicht | Unmöglich | Unmöglich |
| Einfluss von Schmutz und Nässe | Seht stark | Sehr stark | Möglich (Kontakt) | Keinen Einfluss |
| Einfluss von (opt.) Abdeckung | Totaler Ausfall | Totaler Ausfall | Möglich | Keinen Einfluss |
| Einfluss von Richtung und Lage | Gering | Gering | Sehr hoch (nur eine Steckrichtung) | Keinen Einfluss |
| Abnutzung / Verschleiss | Bedingt | Bedingt | Bedingt | Keinen Einfluss |
| Anschaffungskosten / Leseelektronik | Sehr gering | Mittel | Gering | Mittel |
| Unbefugtes kopieren / ändern | Leicht | Leicht | Schwierig | Schwierig |
| Lesegeschwindigkeit (inkl. Handhabung des Datenträgers) | Gering $\tilde{4}$ s | Gering $\tilde{3}$ s | Gering $\tilde{4}$ s | Sehr schnell $\tilde{0}.5$ s |
| Maximale Entfernung zwischen Datenträger und Lesegerät | 0 - 50 cm | $< 1$ cm (Scanner) | Direkter Kontakt | 0 - 100 m, Mikrowelle |

### Kombination von aktiven und passiven RFID Systemen

Um die Kombination von aktiven und passiven RFID Systemen zu beschreiben, soll als Beispielanwendung ein Lokalisierungssystem von Schiffscontainern dienen. Weitere Anwendungsbeispiele werden im Anwendungsteil dieser Seminararbeit vorgestellt.
Das Ziel von RFID Systemen in offenen Wertschöpfungsketten ist, immer sehen zu können, welche Einheiten sich gerade wo befinden. Die vielen kleinen Einheiten, die mit passiven RFID Tags versehen sind und sich im Schiffscontainer befinden, können nicht global lokalisiert werden. Der Grund dafür ist, dass der Schiffscontainer aus Metall ist und die RFID Tags von den Lesegeräten somit nicht gelesen werden können. Ein anderes Problem könnte sein, dass die Lesegeschwindigkeit der verwendeten Lesegeräte zu klein ist, um die vielen passiven RFID Tags aufs Mal auszulesen.
Mit der Kombination von aktiven und passiven RFID Systemen soll dieses Problem gelöst werden. Dafür wird ein schon mit einem aktiven RFID Tag ausgerüsteter Schiffscontainer

benutzt. Die in diesen Container eingeladenen Einheiten, welche mit passiven RFID Tags ausgerüstet sind, können vom aktiven Tag abgefragt werden. Passiert nun der Container ein Lesegerät, wird der aktive RFID Tag gelesen. Er teilt dem Lesegerät auch die in ihm enthaltenen Einheiten mit, welche somit geografisch genau lokalisiert werden können.

### 3.2.5 RFID Standards

Es existieren sehr viele verschiedene Standards zur RFID Technologie [11]. Diese Standards kommen unter anderem vom ANSI (American National Standards Institute), vom Auto-ID Center, vom EAN.UCC (European Article Numbering, Uniform Code Council), vom ERO (European Radio Communications Office), vom ETSI (European Telecommunications Standards Institute) und von der ISO (International Standards Organisation). Im Folgenden wird nur auf die wichtigeren Standards eingegangen, den ISO Standards und den Auto-ID Center Standards.

Standards sind ein sehr wichtiger Faktor wenn es um die Interoperabilität der RFID Technologie geht. Es existieren sehr viele verschiedene Hersteller von RFID Tags und Lesegeräten, die alle untereinander kommunizieren können sollten. Daher müssen gemeinsame Schnittstellenstandards geschaffen werden.

Die Standards können in folgende Gruppen eingeteilt werden [10]:

1. Air Interface Protocol
   Diese Protokolle beschreiben, wie sich der RFID Tag und die Lesegeräte untereinander verständigen.
2. Übermittelte Daten
   Wie sind die Daten auf dem RFID Tag organisiert und formatiert.
3. Konformität
   Diese Standards sind dazu da, zu testen, ob ein Produkt bestimmte Standards erfüllt.
4. Anwendungen
   Diese Standards regeln zum Beispiel wie die Beschriftung von Versandadressen vorgenommen wird.

In den nächsten zwei Abschnitten werden zuerst die ISO Standards behandelt und anschliessend die Auto-ID Center Standards.

#### ISO Standards

Die Internationale Organisation für Standardisierung (ISO) hat verschiedene Standards definiert. Im Folgenden werden sie in die oben genannten vier Gruppen eingeteilt.

Zur ersten Gruppe hat die ISO den Standard 11785 eingesetzt in der Kuherkennung, den Standard 14443, der auf Smartcards eingesetzt wird und den Standard 15693 für so

genannte „vicinity cards“ geschaffen. Die ISO hat desweiteren verschiedene Standards für die Produkteerkennung und die Produkteverwaltung erstellt. Sie sind unter der 18000 Serie zusammengefasst. Hier gibt es eine Klassifizierung nach den beanspruchten Frequenzen des Air Interface Protokolls wie Tabelle 3.2 zeigt.

Tabelle 3.2: Die ISO 18000 Serie

| | |
|---|---|
| 18000-1 | Allgemeine Angaben zu den global akzeptierten Air Interface Frequenzen |
| 18000-2 | Air Interface für 135 kHz |
| 18000-3 | Air Interface für 13.56 MHz |
| 18000-4 | Air Interface für 2.45 GHz |
| 18000-5 | Air Interface für 5.8 GHz |
| 18000-6 | Air Interface für 860 bis 930 MHz |
| 18000-7 | Air Interface für 433.92 MHz |

In der zweiten Gruppe den Standard 11784 in der Kuherkennung. Er regelt genau was für Daten wie zum Beispiel ein Ländercode gespeichert werden [14].

In der dritten Gruppe den Standard 18047 um die Konformität von RFID Tags und Lesegeräten zu testen. In diese Gruppe fällt auch der Standard 18046 der die Performance von RFID Tags und Lesegeräten testet.

In der vierten Gruppe hat die ISO Standards definiert, um grosse Versandcontainer, Palletten, Transporteinheiten oder einzelne Produkte über die Wertschöpfungskette zu verfolgen. Hier sei der Standard CEN/TC 278 erwähnt. Er betrifft verschiedene Aspekte des Transportes von Produkten auf der Strasse [12]. Diese Standards befinden sich in verschiedenen Prüfungsphasen.

### Auto-ID Center Standards

Was die ganze Standardisierung nicht gerade einfach macht, ist die Tatsache, dass das Auto-ID Center seine eigenen Standards für das Air Interface Protokoll entwickelte. Das RFID System vom Auto-ID Center sollte global sein und auf offenen Standards aufbauen. Das Air Interface Protocol vom Auto-ID Center wurde an EPCglobal lizenziert mit der Bedingung, es gebührenfrei für Hersteller und Endkunden verfügbar zu machen.
Das Auto-ID Center entwickelte auch Standards für die Netzwerkarchitektur, welche verschieden von den ISO Standards sind. Diese erlaubt es jedermann Informationen mittels der Seriennummer auf dem RFID Tag über das Internet zu erhalten (siehe EPC). Der Grund für die eigenen Standards des Auto-ID Centers waren die zu hohe Komplexität der ISO Standards. Dies hätte die Kosten für RFID Tags unnötig hoch getrieben. Das Auto-ID Center hat ursprünglich die RFID Tags in sechs Klassen eingeteilt:

- Klasse 1
  Einfacher, passiver RFID Tag, der nur gelesen werden kann. Er ist genau einmal beschreibbar.

- Klasse 2
  Passiver RFID Tag mit bis zu 65 KB Speicher, der gelesen und in den geschrieben werden kann.

- Klasse 3
  Halbpassiver RFID Tag, d.h. ein Klasse 2 RFID Tag der aber zusätzlich mit einer Batterie ausgestattet ist, um auf weitere Distanz gelesen zu werden.

- Klasse 4
  Aktiver RFID Tag, auch mit einer Batterie ausgestattet, um das Signal zu einem Lesegerät zu übertragen.

- Klasse 5
  Aktiver RFID Tag, der mit anderen Klasse 5 RFID Tags oder anderen Geräten (zum Beispiel Lesegerät) kommunizieren kann.

Hier kommt noch eine weitere Klasse hinzu, nämlich die Klasse 0. Sie ist insofern speziell, als dass sie ein anderes Air Interface Protokoll benutzt als die Klassen 1 bis 5. Der Klasse 0 RFID Tag kann auch nur gelesen werden und wird unmittelbar bei der Produktion beschrieben.
Diese Einteilung hat sich aber über die Zeit ein wenig geändert. Die Gründe waren die Anzahl der Leseoperationen und die Distanz.
Die Klasse 0 und Klasse 1 Standards des Auto-ID Center haben zwei Schwächen. Erstens sind sie nicht kompatibel mit den ISO Standards. Sie könnten bei der ISO angemeldet werden, würden aber nicht akzeptiert, weil sie nicht den ISO Standards entsprechen. Zweitens können Klasse 0 RFID Tags nicht global verwendet werden, weil sie auf einer Frequenz Signale erhalten und auf einer anderen zurücksenden. Dies ist zum Beispiel in Europa nicht erlaubt.
Im Jahr 2004 wurde vom Auto-ID Center ein zweitgeneration Protokoll (Gen 2) entwickelt, das nicht rückwärtskompatibel zu den Klasse 0 und Klasse 1 Standards war. Damit sollte ein globaler Standard geschaffen werden, der näher an den ISO Standards des Air Interface Protokolls lag.

Bis jetzt sind die beiden Organisationen noch zu keinen gemeinsamen Standards gekommen. Das ist jedoch für eine weitere Verbreitung der RFID Technologie notwendig. Es wird damit gerechnet, dass dieser Annäherungsprozess noch mehr als ein Jahr andauern wird [10]. Ferner ist nicht bekannt wie sich grosse Industrienationen, wie zum Beispiel China, aufgrund fehlender Standards verhalten werden [4].

### 3.2.6 EPC (Electronic Product Code)

Die Abkürzung EPC wird auf zwei Arten verwendet [15]. Vorrangig beschreibt diese Abkürzung einen Zahlencode, den Electronic Product Code. Ganz allgemein kann der Ausdruck auch zusammenfassend für das gesamte EPC Netzwerk verwendet werden wie es in Abbildung 3.4 dargestellt ist [18]. Die wesentlichen Komponenten sollen hier erläutert werden. Auf der linken Seite sind die Funktionen dargestellt, die direkt mit dem RFID

Tag zusammenhängen. Sie werden hier nicht näher beschrieben. Auf der rechten Seite befinden sich die Systeme, die mit Hilfe des EPC und ONS Informationen zu einem Produkt aus dem Internet beziehen können. Der ONS wird weiter unten näher beschrieben.

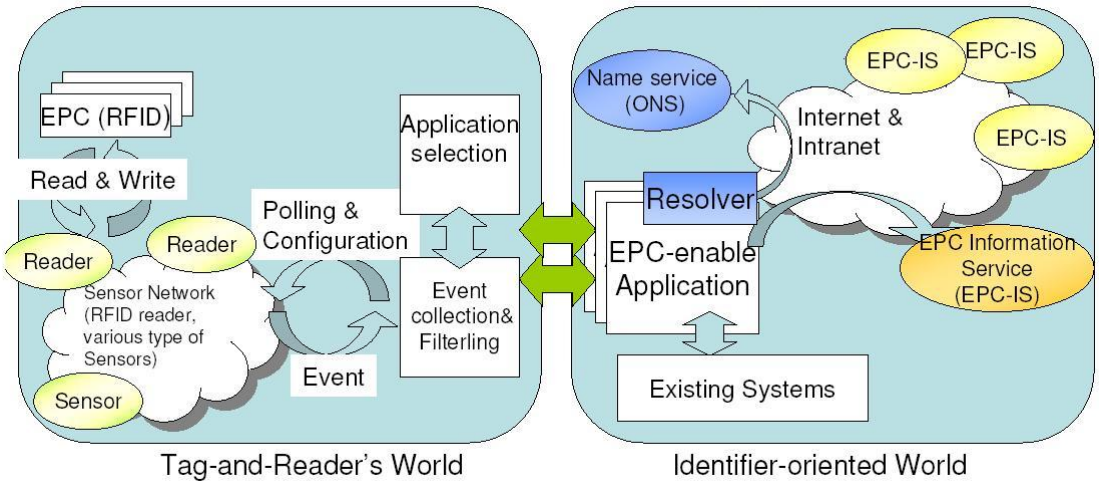


Abbildung 3.4: Übersicht EPC Netzwerk

Es wurde bei der Gründung des Auto-ID Center im Jahr 1999 begonnen, EPC zu entwickeln. Mit Hilfe des EPC sollte es möglich sein, Produkte zu identifizieren und über die ganze Wertschöpfungskette global zu verfolgen. Ein Beispiel aus dem Anwendungsteil sei hier erwähnt: Hühnereier werden so über die gesamte Wertschöpfungskette verfolgt.

Der EPC ist eine Ziffernfolge, bestehend aus einem so genannten Header und drei weiteren unabhängigen Abschnitten zur Speicherung von Informationen. Es sind bisher drei Grundversionen des EPC entwickelt worden.
Die ursprüngliche Version des EPC hat eine Länge von 96 bit (EPC-96). Er war jedoch zu gross um kostengünstig auf den RFID Tags gespeichert und durch die eingesetzte Hardware verarbeitet zu werden. Daher soll kurzfristig ein 64 bit (EPC-64) langer EPC zum Einsatz kommen. Durch seine geringe Länge soll Speicherplatz gespart werden, was den Einsatz kostengünstigerer RFID Tags erlaubt [16].
Die letzte Ausbaustufe wird ein 256 bit langer EPC sein. Nach aktuellen Untersuchungen besteht jedoch derzeit kein dringender Bedarf für einen 256 bit EPC, da der EPC-96 für heutige Verhältnisse schon genug Spielraum bietet. Eine Veranschaulichung wie die Bits für den EPC-96 aufgeteilt sind bietet Abbildung 3.5 [16].

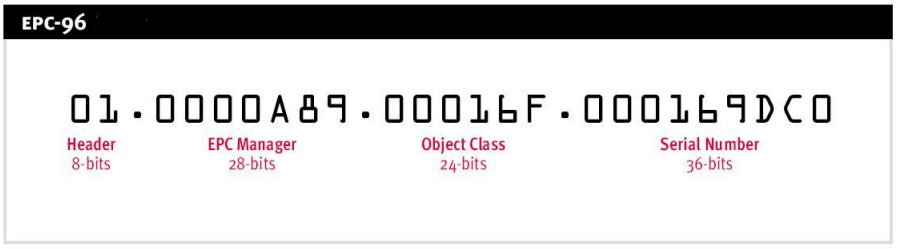


Abbildung 3.5: Bitaufteilung beim EPC-96

Wichtig ist in diesem Zusammenhang, dass die verschiedenen EPC Versionen kompatibel sind. Es wird also jeder Benutzer auf eine höhere Version umsteigen können, ohne die gesamte Software austauschen zu müssen.

Der grundsätzliche Aufbau der verschiedenen EPC Versionen ist gleich. Der Header dient zur Kennzeichnung der EPC Version. Diese Versionsnummer beinhaltet die Anzahl, den Typ und die Länge aller nachfolgenden Datenelemente. Der nächste Teil des EPC heisst EPC-Manager und stellt die Kennzeichnungsnummer des Inverkehrbringers, in der Regel des Industrieunternehmens, dar. Der dritte Abschnitt des EPC nennt sich Object Class. Hier erscheint die Nummer des Produkts. Im letzten Feld ist eine Seriennummer untergebracht. EPC-Manager und Object Class entsprechen bei der Identifikation von Artikeln der im EAN.UCC System verwendeten EAN. EAN steht für die Europäische Artikel Nummer und besteht meistens aus 13 Ziffern. Bekannt ist die EAN vom Strichcode, der sich zum Beispiel auf jedem Lebensmittel befindet. Die EAN besteht aus einem ersten Teil, der dem EPC-Manager entspricht und einem zweiten Teil, welcher der Object Class entspricht [17]. Somit ist der EPC eine EAN, die um eine Seriennummer erweitert wird. Es sind aber auch andere Objekte wie zum Beispiel logistische Einheiten wie Paletten mit einer EPC identifizierbar.

### ONS (Object Name Service)

Der ONS (Object Name Service) ist ein Dienst, der das Auffinden von Produktinformationen anhand eines EPC über das Internet ermöglicht. Diese Vorgehensweise erlaubt es, nur wenig Informationen auf dem RFID Tag zu speichern (nur den EPC), was einen günstigeren RFID Tag möglich macht. Eines der zwei Ziele des Auto-ID Centers ist ja, einen günstigen RFID Tag herzustellen.
Der ONS ist mit dem DNS (Domain Name Service) aus dem Internet vergleichbar. Die Funktionsweise ist Folgende: Der ONS erhält über einen Resolver eine Anfrage zu einer EPC. Anhand des EPC wird eine passende URL oder IP-Adresse aus der Datenbank gesucht, die zurück gesendet wird. So kann mit der erhaltenen Adresse auf die Produktinformation zugegriffen werden (ein EPC Information Service). Eine Veranschaulichung hierfür liefert die Abbildung 3.4.
Der ONS benützt als Kommunikationsprotokoll das DNS Protokoll. Er ist hierarchisch aufgebaut mit einem Rootserver. Dieser zeigt auf die darunterliegenden EPC-Manager Server, die von den Inverkehrbringern betrieben werden. Diese wiederum zeigen auf die EPC Information Service Einträge, die sich auf die Object Class des EPC beziehen. Unter diesen Einträgen ist wie oben beschrieben die Produkteinformation verfügbar [18].

Der ONS ist, wie der DNS, darauf ausgelegt, die Adressinformation in einem Verbund von ONS-Servern verteilt zu speichern. Es besteht die Möglichkeit, bereits abgerufene EPC-IP Paare lokal zu speichern. Die lokale Speicherung hat den Effekt, dass das ONS-Netzwerk nicht durch oftmaliges Abfragen derselben Information unnötig belastet wird. Die abgerufenen Informationen können auch lokal gespeichert werden, wodurch das Netzwerk weiter entlastet wird und der Kunde schneller an die gewünschte Information gelangt.

**PML (Physical Markup Language)**

Da mit EPC viele Informationen über ein bestimmtes Produkt über das Internet abgefragt werden können, sollten diese Informationen auch standardisiert ausgedrückt werden können. Dafür verwendet man die PML, die vom Auto-ID Center entwickelt worden ist. Die PML ist eine computerorientierte Sprache zur Beschreibung von physischen Objekten. Statt eine völlig neue Sprache zu erfinden, ist bei der Entwicklung des EPC-Konzepts von einer bereits existierenden Sprache ausgegangen worden, nämlich XML. PML beschreibt, ähnlich wie XML, wie Daten dargestellt werden (zum Beispiel ein Text wird zentriert, linksbündig, dargestellt) und auch welche Daten beschrieben werden (Datum, Zeit, ...). PML unterstützt verschiedene Datentypen zur Darstellung dieser Information. PML wird also für die elektronische Darstellung von Objekten und für den Transport dieser Information verwendet [19].

## 3.3 RFID Anwendungen

Wie im Kapitel Einführung in RFID beschrieben wurde, beeinflusst die Leistungssteigerung, welche durch RFID ermöglicht werden, sowie auch die günstigen Produktionsverfahren der Prozessoren beeinflussen zahlreiche Anwendungsgebiete. RFID kann auch als Fundament für weitere Anwendungen in betracht gezogen werden. Die Anwendungsbereiche von RFID erstrecken sich beinahe über jeden Wirtschaftszweig.
RFID wird derzeit im Einzelhandel stark diskutiert. Grosse Handelskonzerne, wie Tesco (Grossbritannien) [21], Metro Group (Deutschland) [22] und Wal-Mart (USA) [23], gehören zu den Treibern dieser Technologie in ihrer Branche. Neben dem Einzelhandel ist auch die Logistikbranche RFID interessiert. Die Steigerung der Effizienz und die Automatisierung von Abläufen bergen für diese Branche grosse Einsparungspotentiale.

### 3.3.1 Anwendungen für RFID

Wie schon angesprochen sind die Anwendungsgebiete von RFID über verschiedene Wirtschaftssegmente ausgedehnt. Die folgenden Anwendungsgebiete können jedoch Branchen übergreifend unterschieden werden [24].

- Kennzeichnung von Tieren und Objekten
- Echtheitsprüfung von Dokumenten
- Instandhaltung und Reparatur, Rückrufaktionen
- Diebstahlsicherung und Reduktion von Verlustmengen
- Zutritts- und Routenkontrolle
- Supply Chain Management

### 3.3.2 Kennzeichnung von Tieren und Objekten

Durch die grössere Speicherkapazität, sowie durch den nicht benötigten Sichtkontakt bzw. die kontaktlose Erkennung von RFID ist die Kennzeichnung von Tieren und Objekten ein relevantes Einsatzgebiet für RFID Technologie. Bereiche wie die Nutz- und Heimtieridentifikation, die Transportidentifikation von Gefahrgüter, aber auch Personenidentifikation. Die Identifikation von Nutztieren gibt es schon länger. Die gelben Ohrmarken, wie man sie bei Kühen auf den Weiden oder bei Schweinen im Stall sieht, vgl. Abbildung 3.6, sind den meisten Leuten ein begriff. Die schmerzhafte Markierung von Nutztieren mittels eines Brandeisens oder das tätowieren von Tieren gehören ebenso zu den bekannten Identifizierungsmassnahmen.


Abbildung 3.6: Traditionelle Ohrmarken

Mittels RFID können Nutztiere nun auf elektronischem Weg identifiziert werden. Dabei ist dieses Verfahren schneller und auch sicherer gegen Betrug. Um Nutztiere zu identifizieren existieren drei verschiedene Möglichkeiten. Die elektronische Ohrmarke, ein Transponder-Injektat welches auch bei Haustieren wie Hund und Katze eingesetzt werden kann und der Bolus. Beim Bolus handelt es sich um einen Keramikzylinder, vgl. Abbildung 3.7, indem ein in Silikon versiegelter RFID Transponder enthalten ist. Der Bolus wird dem Tier oral, vgl. Abbildung 3.8 [25], verabreicht und bleibt anschiessend im Vormagentrakt dem sog. Pansen, d.h. der Bolus wird nicht wieder ausgeschieden.

Durch die schnelle und lückenlose Identifizierung jedes einzelnen Nutztieres kann das Tier von seiner Geburt bis zur Schlachtung und damit zum verkauf des Fleisches verfolgt werden. Nach den Skandalen wie BSE in ganz Europa und der Maul- und Klauenseuche in Grossbritannien, wird von den Behörden und den Konsumenten eine Verbesserung der Verfolgung und Herkunft der Tiere gefordert.

Bei der Kennzeichnung von Tieren werden passive Tags eingesetzt. Die Auslesedistanz kann bis 130 Zentimeter betragen, was auf High Frequency Tag schliessen lässt. Der Speicher des Tags muss nicht sehr gross sein, weil spezifische Informationen zu den Tieren zusätzlich in einer Datenbank gespeichert werden können.
Die Identifikation mit RFID ist nicht nur für Nutztiere vorgesehen. Der Bundesrat hat im


Abbildung 3.7: Bolus: RFID versiegelt in Keramikzylinder

Abbildung 3.8: Verabreichung von Bolus an Nutztiere

Juli 2004 beschlossen, ab Januar 2006 alle Welpen mit einem RFID Transponder zu markieren [26]. Die Europäische Union verlangt schon heute bei der Einreise von Hunden, dass diese durch einen Transponder eindeutig identifizierbar sind. Zudem erleichtert die Einführung einer Transponder Implantation die Zuordnung von Hunden die in Beissunfälle verwickelt sind, sowie entlaufene oder ausgesetzte Tiere.

Wie oben bereits angesprochen kann die RFID Technologie auch für die Markierung von Gefahrgütern verwendet werden. Durch die, im Vergleich zu Barcodes, höhere Speicherkapazität, die Unempfindlichkeit gegen Schmutz und anderen Umwelteinflüssen, ist RFID auch hier im Vorteil. Neben den passiv gespeicherten Daten welche ein Transponder enthalten kann, können sich beschreibbare Transponder auch aktualisieren lassen. Dadurch ist es möglich auf veränderte Situationen einzugehen.

Ein weiteres Wichtiges Anwendungsgebiet, das hier nur kurz behandelt wird ist im Gesundheitswesen. Aktive Transponder an Blutkonserven können auf Temperaturveränderungen aufmerksam machen und so Patienten vor verdorbenen Blutkonserven schützen. Blinden Menschen kann die RFID Technologie das Suchen von Objekten erleichtern. Objekte die mit einem RFID Tag ausgestattet sind könnten von einem tragbaren Terminal einfach gefunden werden.

### 3.3.3 Echtheitsprüfung von Dokumenten

Durch die Angst vor Terroranschlägen verlangen Staaten, allen voran die USA, Reisedokumente mit Biometrischen Merkmalen. Auf der diesjährigen CeBIT (10. - 16. März 2005) stellte die deutsche Bundesdruckerei den mittels RFID digitalisierten Reisepass vor [27]. Auf den können biometrische Daten digital gespeichert werden. An Kontrollstationen, wie zum Beispiel bei der einreise in ein fremdes Land, wird der Pass dann gescannt, d.h. der Transponder wird ausgelesen und die Person kann aufgefordert werden Finger oder Gesicht mittels eines Verifier überprüfen zu lassen. Heise Online zufolge seien die ersten Tests mit dem neuen Reisedokument erfolgreich verlaufen, Probleme entständen nur, wenn Visa eingeklebt würden. Es muss auch noch abgewartet werden, ob der Transponder die jahrelange Verwendung übersteht.

Über die technischen Eigenschaften des zu verwendenden Tags kann bis zum jetzigen Zeitpunkt, der Internetseite der deutschen Bundesdruckerei, keine Informationen entnommen werden. Es ist jedoch anzunehmen, dass es sich um einen passiven Tag mit Low Frequency handeln wird. Die Auslesedistanz muss reduziert sein, um den Datenschutz der einzelnen Personen möglichst zu gewährleisten. Der Internationale Standard sieht einen Speicher von 32 KB vor. Damit lassen sich die Rohdaten für ein biometrisches Merkmal speichern. Länder wie die USA verlangen jedoch die Speicherung zweier biometrischen Merkmalen, die auf einem 64 KB Tag gespeichert werden müssen.
Eine Feuertaufe steht der RFID Technologie im Jahre 2006 bevor [28] [29] [30]. Die Eintrittskarten an der Fussball Weltmeisterschaft werden auch mit RFID Tag ausgerüstet sein. Obwohl diese Ausführung auch weiter unten in „Zutritts- und Routenkontrollen“ besprochen werden kann, werde ich hier darauf eingehen. Die Veranstalter der Fussball Weltmeisterschaft wollen durch die elektronische Zugangstechnik das Fälschen von Eintrittskarten erschweren. Jede Eintrittskarte ist Personengebunden, das heisst nur die Person, die das Ticket erworben hat ist berechtigt das Stadion zu betreten. Soll eine erworbene Karte weiterverkauft werden, so muss der Käufer die Zustimmung des FIFA Fussball - Weltmeisterschafts- Organisationskomitees Deutschlands einholen. Damit kann das OK gewährleisten, dass nur von ihnen berechtigte Personen Zutritt erhalten. Bekannte Hooligans können so von der Weltmeisterschaft 2006 ferngehalten werden. Auch auf den Tickets werden passive Tag verwendet, welche im Low Frequency Bereich arbeiten.

### 3.3.4 Wartung und Reparatur

Seit der Brandkatastrophe im Düsseldorfer Flughafen, ist bekannt dass die Wartung der Brandschutzklappen der Lüftungsanlagen eine sehr wichtige Rolle spielt. Am Münchner Flughafen mussten früher Techniker zu den Lüftungsklappen steigen, diese überprüfen und noch einen dreiseitigen Bericht auf Papier erfassen. Auch hier hat kam die RFID Technologie zum Einsatz [31]. Jede Klappe wurde mit einem passiven Tag ausgerüstet. Die Techniker wurden mit einem Psion Handheld ausgerüstet, der in der Lage ist die Tags auszulesen. Um Schlamperei vorzubeugen wurde die Reichweite der Tags auf drei Zentimeter reduziert, so dass die Techniker gezwungen sind die Klappen auch wirklich aufzusuchen. Die Firma SAP die auch an neuen Geschäftsprozessen interessiert war, stellte für die Handhelds einen R/3-Client zu Verfügung, um Information und Datum der letzten

Wartung zu speichern. Die Folgen der Einführung der RFID Technologie sind folgende. Der Monteur muss keine Papiere mehr ausfüllen, und die Daten sind direkt weiterverwendbar. Letzteres gilt nicht nur für die normalen Instandhaltungsprozesse, sondern auch für Auswertungen im R/3-Reporting oder im „SAP Business Warehouse“. Sind erst einmal genug Informationen aufgelaufen, lassen sich häufig auftretende Defekte erkennen oder die Lebensdauer und Störanfälligkeit unterschiedlicher Produkte miteinander vergleichen [31].

Was beim Reifenhersteller Michelin durch logistische Überlegungen begann, soll bald noch weitere Wellen schlagen. Michelin will in für logistische Zwecke ihre Reifen mit Tags ausrüsten, um ihre logistischen Prozesse zu optimieren [32]. Die Tags können aber auch gleich dazu verwendet werden, den Fahrer über gefährlichen Unterdruck der Reifen zu informieren. Dies soll folgendermassen funktionieren. Ein Sensor überwacht die Rotationsfrequenz des Reifen. Ein Reifen mit Unterdruck sind im Durchmesser etwas kleiner und rotieren dementsprechend schneller. Dies kann erkannt und dem Fahrer mitgeteilt werden. Das ist das sog. indirekte Messsystem. Das direkte System kann durch direkte Überwachung des Reifens den Reifendruck erkennen und potentiell gefährliche Druckabweichungen melden.

### 3.3.5 Zutritts und Routenkontrolle

Nicht nur in IT Bereichen müssen sich Personen durch Passwörter und Benutzernamen ausweisen, um Zugang zu bestimmten bereichen zu erhalten. Auch im alltäglichen Leben und in der Berufswelt werden immer mehr Zutrittskontrollen aufgebaut. Neben dem normalen Schlüssel finden vermehrt auch Magnetkarten und Chipkarten ihre Anwendung. Viele der heute im umlaufbefindlichen Karten müssen am Kontrollpunkt in ein dafür vorgesehenes Terminal eingeführt werden. Die modernen elektronischen Ausweise funktionieren anders. Die zu überprüfende Person kann muss die Erfassungsgeräten mit einem maximalen Abstand von ca. einem Meter passieren. In der Abbildung 3.9 ist eine Erfassungsstation der Firma Nagra zu sehen [33].

Abbildung 3.9: Skilift mit Handfreetechnologie

Solche Stationen sind zum Beispiel in der Weissen Arena in Flims Laax Falera installiert. Diese als „Handfree Technologie“ bezeichneten Systeme erlauben es den Besuchern die Drehkreuze der Skilifte zu passieren, ohne dabei die Karte hervor zu suchen und in das

Erfassungsgerät einzuführen. Damit ist es jedoch noch nicht getan. Es existieren schon Projekte wobei eine Ressortlösung angestrebt wird [34]. Den Gästen der Weissen Arena soll mittels einer einzigen SmartCard die Nutzung sämtlicher Dienstleistungen des Ressorts ermöglicht werden. Transportunternehmen Sportanlagen, Hotels Restaurants, Shops, Vermietungsfirmen, etc. können in dieses Netzwerk eingebunden werden. Ein Gast braucht also nur noch die SmartCard mit sich zu führen um sich auszuweisen, ihr Hotelzimmer zu öffnen oder zum Beispiel das Mittagessen oder den Drink im Ausgang bezahlen zu können. Die Betreiber erhoffen sich eine hohe Akzeptanz der Kunden. Sie setzen dabei vor allem auf die Bequemlichkeit. An die Karte sollen darüber hinaus auch noch ein Kundenbindungsprogramm gekoppelt werden. Kunden haben so die Möglichkeit mit der Karte Punkte zu sammeln um so wieder von Angeboten profitieren zu können. Für die Anbieter soll die Karte zu einer Optimierung der Geschäftsprozesse führen. Sie sollen von Synergien profitieren können, welche durch die gemeinsame Vermarktung und der professionellen Kundenpflegen entstehen sollen.

In anderen Klimaregionen, genauer in Südafrika, laufen Versuchen, Mautabgaben mittels RFID einzuführen [35]. Die Nummernschilder der Testfahrzeuge wurden mit passiven RFID-Tags ausgestattet. Die Empfänger durften höchstens einen Abstand von 6 Meter zu den zu überprüfenden Fahrzeugen aufweisen, was auf High bis Ultra High Frequency schliessen lässt. Die Lesegeräte waren imstande bis zu 7200 Fahrzeuge pro Minute auszulesen, dies ist auch bei starkem Verkehrsaufkommen durchaus ausreichend. Dabei konnten die Fahrzeuge mit Geschwindigkeiten von bis zu 250 Stundenkilometern durch die Lesestationen rasen, ohne dass dabei die Informationen verloren gingen. Die Lebensdauer der Tags betrug bis zu sieben Jahren, wobei sie Temperaturschwankungen von -40 Grad bis +85 Grad Celsius auszuhalten hatten.

Die Routenkontrolle ist auch beim Sport wichtig. Am „Züri Marathon“ der kürzlich stattgefunden hat trug jeder Läufer an seinen Schnürsenkeln festgemacht einen passiven Transponder. Die Auslesedistanz ist relative gering. Die Läufer müssen über eine Matte laufen, unter welcher sich die Auslesestation befindet. Es ist daher anzunehmen, dass der Transponder im Low Frequency Bereich arbeitet. Einen solchen Transponder ist auf der Abbildung 3.10 zu sehen.

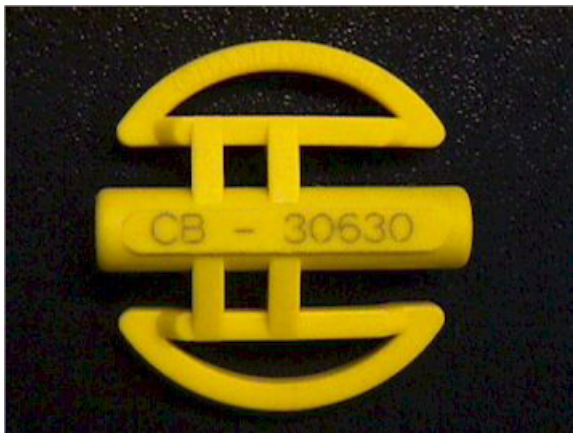


Abbildung 3.10: RFID Chip für Laufveranstaltungen

Durch diesen Transponder können die Organisatoren feststellen, ob ein Läufer vorgegebene

Punkte wirklich durchlaufen hat, beim „Züri Marathon“ waren dies zum Beispiel der Wendepunkt in Meilen sowie natürlich auch Start und Ziel. Die Zeitnahme wird ohne jedes aktive Zutun der Sportler durchgeführt. Bei der Überquerung der Lesestation werden die Daten ausgelesen. So ist es jedem Sportler möglich seine individuelle Start und Endzeit auszulösen. Nach der Ankunft im Ziel wird zudem gleich die Rangliste erstellt und die Daten sind danach im Internet abrufbar. Das System ist nach Angaben der betreibenden Firma in der Lage je nach Konfiguration bis zu 1000 Sportler in der Minute zu erfassen. Diese Dienstleistung wird von der Firma Datasport bei vielen Sportveranstaltungen in der Schweiz verwendet [36].

### 3.3.6 Diebstahlsicherung und Reduktion von Verlustmengen

Um Autobesitzer vor dem Diebstahl ihrer Fahrzeuge zu beschützen haben Hersteller wie beispielsweise Ford, Nissan, Toyota ihre Fahrzeuge mit einer Wegfahrsperre ausgerüstet. Integriert im Zündschlüssel der Fahrzeuge befindet sich ein RFID Tag, welcher die Wegfahrsperre ausschaltet. Ist der Tag nicht vorhanden, so wird der Zündkontakt nicht hergestellt, ein Kurzschliessen der Fahrzeuge soll ebenfalls nicht möglich sein.

Wer an Diebstahlsicherung denkt, dem fällt vielleicht nicht als erstes die Wegfahrsperre von Kraftfahrzeugen, sondern eher die Lesegeräte, welche in den Kaufhäusern und Einzelhändler bei den Ausgängen stehen ein. Auch hier wird die RFID Technik eingesetzt um Ladendiebe von ihrem Vorhaben abzubringen [37]. Auf der Abbildung 3.11 ist ein solches Lesegerät mit schematischen Funkwellen zu sehen. Die passiven Transponder, welche meistens auf der Etikette des Produktes befestigt sind müssen beim bezahlen an der Kasse deaktiviert werden, damit sie den Alarm beim verlassen des Geschäftes nicht auslösen. Um sich vor den Abschneiden oder Austauschen der Etiketten zu schützen, könnten die RFID Tags auch in die Produkte mit eingearbeitet werden, so dass sie nicht vom Produkt separiert werden können. Solche Lösungen können aber natürlich auch über die Diebstahlsicherung weiter verwendet werden. Wie bereits ausführlich besprochen wurde können diese Tags auch Informationen speichern, welche auch von Kunden am Point of Sale benutzt werden können. RFID Tags an Weinflaschen können über dessen Herkunft informieren. Vorschläge zu Gerichten und ganze Menütipps könnten auch bereitgestellt werden. Der Phantasie sind in diesem Bereich wohl kaum Grenzen gesetzt.

### 3.3.7 Supply Chain Management

Supply Chain Management (Wertschöpfungskette) ist ein Unternehmen übergreifender Prozess. Es geht dabei darum, aus Rohstoffen wie zum Beispiel Bäumen, Nutzvieh und Eisenerz, ein Endprodukt wie zum Beispiel ein Sofa zu erzeugen, welches durch Endkunden gekauft werden kann. Je effizienter diese Prozesse gesteuert wird desto kostengünstiger kann produziert werden. In der optimalen Supply Chain werden die Lager sozusagen aufgehoben. Es wird alles Just in time geliefert und produziert. Ein im Liefereingang ankommendes Produkt wird gleich in die Produktion überführt. Diese Art von Produktion birgt auch sehr grosse Gefahren. Entsteht nur durch einen Lieferanten einen Lieferengpass,

Abbildung 3.11: System für Diebstahlsicherung

so kann dadurch die ganze Produktion gestoppt werden, obwohl vielleicht nur ein kleiner Baustein der Produktion nicht vorhanden ist.

Die grosse Vision ist, dass die Wertschöpfungsketten umgekehrt werden können. Heute werden die Produkte meist aufgrund von Vorhersagen durch die Wertschöpfungskette „gestossen“. Da kann es passieren, dass bestimmte Produkte nicht verkauft werden können, weil eine Nachfrage fehlt. In Zukunft will man die RFID Technologie nutzen, um Produkte durch die Wertschöpfungskette zu „ziehen“. Der Verkauf eines bestimmten Produkts kann in Echtzeit verfolgt werden und dem Hersteller direkt mitgeteilt werden. Hier schlägt die grosse Stunde der RFID Technologie. Sie ermöglicht es Produkte und Materialien in Echtzeit bis zum einzelnen Produkt über das ganze Logistiknetzwerk zu verfolgen und frühzeitig evt. Lieferengpässe erkennen und darauf reagieren zu können. Die RFID Tags speichern Daten über Produkte und geben deren zeitlichen und räumlichen Bewegungen wieder. Zudem entstehen keine Medienbrüchen mehr, d.h. Lieferscheine müssen nicht mehr ausgedruckt und später wieder eingelesen werden, auch die Bezahlung kann damit automatisiert werden. Wenn die Ware im Wareneingang angenommen wird, kann das System automatisch den Zahlungsauftrag auslösen.

Die Autoindustrie ist eine typische Just in time Branche. Da in der Automobilindustrie Auftrags gebunden produziert wird und die meisten bestellten Automobile nicht identisch sind, ist die Just in time Produktion enorm wichtig. Dafür ist die automatische Materialflussüberwachung eine der wichtigsten Voraussetzungen um einen ungestörten Betrieb zu garantieren. Wird die Karosserie bei der Produktion mit einem RFID Tag ausgestattet, kann damit die ganze Produktion dieses Fahrzeuges gesteuert werden [38]. Besonders bei der chaotischen Fertigung am Band ist die zweifelsfreie Identifizierung der zu verbauenden Komponenten unabdingbar. In ihm können alle Attribute des zu bauenden Fahrzeuges gespeichert und wieder ausgelesen werden. Der Tag bleibt jedoch auch nach der Fertigstellung des Kraftfahrzeugs in der Karosserie. Dadurch kann an Servicestationen Informationen über die verwendeten Baugruppen für die Ersatzteilbeschaffung genutzt

werden. Das Bild 3.12 zeigt eine Karosserie eines Mercedes Benz mit dem integrierten RFID Tag.

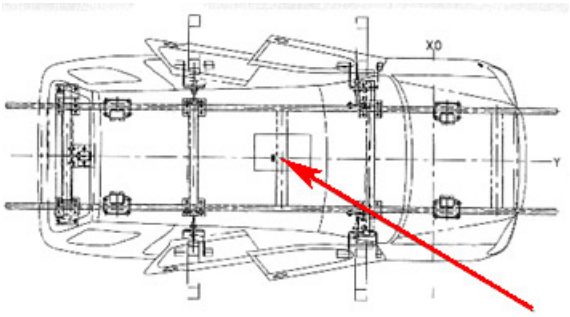

Abbildung 3.12: Autokarosserie mit integriertem RFID Chip

Mit zunehmendem Wettbewerbsintensität und der immer billiger werdenden Technologie, kommt auch zunehmend der Handel in den Genuss der Kostensenkungspotentiale der RFID Technologie. Die Kostensenkungspotentiale ergeben sich durch Reduzierung der Lager und somit kleiner werdenden Kapitalbindungskosten und den Einsparungen an Personal in Lagern und Geschäften. Der bis heute noch nicht übertroffene Höhepunkt ist der durch die Metro Group etablierte Future Store [22]. Der Future Store wurde entwickelt um technische und prozessuale Entwicklungen und Innovationen im Handel zu testen, weiterzuentwickeln und zu etablieren. Die dort eingesetzten Technologien wie RFID machen die Prozesse im Handel schneller, effizienter und transparenter. Bestellvorgänge, Lieferungen und die Lagerung von Waren werden vereinfacht. Darüber hinaus ergeben sich zahlreiche Möglichkeiten, mit zielgruppenorientierten Angeboten auf die Bedürfnisse der Kunden einzugehen. Der Handel profitiert davon gleich zweifach: Die optimierten Prozessabläufe führen zu Kosteneinsparungen und zu mehr Kundenzufriedenheit.
Am Beispiel von Hühnereiern kann das Potential von RFID im Supply Chain Management im Handel verdeutlicht werden, vgl. Abbildung 3.13. Denkbar ist, dass bereits der Erzeuger die Kartons mit Smart Chips versieht. Angaben wie Legedatum, Legebetrieb, Futtermittel und Haltungsform - beispielsweise Freiland- oder Bodenhaltung - werden in einer Datenbank hinterlegt. Die Lieferscheine werden automatisch erstellt und elektronisch an das Zentrallager übermittelt. Dort registrieren Lesegeräte den Wareneingang und gleichen die Zahl der gelieferten Eier mit der bestellten Menge ab. Die ankommenden Chargen können dank RFID problemlos den jeweiligen Zielorten zugeordnet und auf die entsprechenden Lkws verteilt werden. Verlässt eine Lieferung frischer Eier das Zentrallager in Richtung Markt, erhält dieser wiederum einen elektronischen Lieferschein. Auch hier gleichen RFID Schleusen Wareneingang und Bestellung miteinander ab. In den Lagern sind die Gabelstapler und Hochregale ebenfalls mit RFID Technologie ausgestattet - so können die Mitarbeiter Paletten und Kartons gezielt lokalisieren [39].

Trotz Optimierung durch neue Technologien lässt sich die Lagerhaltung nicht immer verhindern. Doch auch hier findet RFID seine Anwendung. In modernen vollautomatisierten Hochregallagern wird die einzulagernde Ware mit Transponder versehen. Da die Transponder nicht fest mit den Waren verbunden werden sondern mittels Halteklammern an den Paletten angebracht werden, können die Transponder über viele Jahre wieder verwendet werden. Die Lieferscheine werden beim Wareneingang gleich elektronisch in die Transpon-


Abbildung 3.13: Supply Chain Management am Beispiel von Hühnereier

der gespeichert, auf diese weise ist es möglich, ohne Sichtkontakt mit der entsprechender Ware, deren Status bis auf die kleinsten Details abzufragen.

## 3.4 Hemmende und Fördernde Faktoren von RFID

Die hemmenden und fördernden Faktoren können sich sowohl auf die Technologie als auch auf die Anwendungsgebiete abstützen. Um die verschiedenen fördernden und hemmenden Faktoren der RFID Technologie aufzuzeigen, ist es wichtig ihre stärksten Konkurrenten kurz zu betrachten und dann die Vor- und Nachteile der verschiedenen Technologien einander gegenüber zu stellen. Die ärgste Konkurrenz Technologie ist zweifelsfrei der Barcode. Weiter sind auch noch OCR (optical character recognition) und die Chipkarten. In der Tabelle 3.1 werden die verschiedenen Parameter der einzelnen Technologien aufgelistet.

### 3.4.1 Fördernde Faktoren

Die Fördernden Faktoren entstehen durch die Vorteile, welche RFID gegenüber den bereits heute etablierten Technologien besitzt.

### Kontaktloses Auslesen

Die Fördernden Faktoren der RFID Technologie liegen vor allem in ihrer Eigenschaft dem kontaktlosen Auslesen. Es weder direkter Kontakt noch direkter Sichtkontakt zwischen der Lesestation und dem RFID Tag benötigt um die Daten, die auf dem Tag gespeichert sind, auslesen zu können. Durch diesen Auslesevorgang ist die Technologie vor Vandalismus schützbar. Anders als zum Beispiel Telefonkabinen, bei denen die Chipkarten durch einen Kartenleser gezogen werden müssen, können die Lesestationen von RFID abgetrennt von den Benutzern aufgestellt werden. Durch das kontaktlose Auslesen ohne Sichtkontakt können auch Produkte im Pulk ausgelesen und verarbeitet werden. Dies ist zum Beispiel bei Barcode nicht möglich, da jeder Barcode Sichtkontakt zum Scanner aufweisen muss, damit die Informationen verarbeitet werden können.

### Robustheit der Technologie

Die Robustheit der Technologie lässt sich zum Teil auch auf das kontaktlose Auslesen zurückführen. Der Verschleiss der bei kontaktbasierten Technologien entfällt. Damit verlängert sich die Lebensdauer der einzelnen Tags. Wiederverwendbare Tags werden dadurch natürlich attraktiver.
Ein zweiter Punkt für die Robustheit der Technologie ist, dass im Vergleich zu ihrer konkurrenzierenden Technologien, die RFID Technologie wenig bis nicht anfällig auf Umwelteinflüsse, wie Schmutz und Nässe ist. Die Barcode Technologie wird, wenn sie verschmutzt wird, unlesbar für das Lesegerät.

### Kosten der Tags

Ein wichtiger Grund dass RFID in nächster Zeit eine wichtige Technologie werden wird, sind die Kosten für RFID Tags, die in der letzten Zeit sehr stark gesunken sind. In der Vergangenheit konnten relativ hohe Kosten für ein RFID Tag (ein US-Dollar oder mehr pro RFID Tag) dadurch gerechtfertigt werden, dass die RFID Tags innerhalb einer Firma blieben und somit wiederverwendet werden konnten. Wenn aber Güter über eine ganze Wertschöpfungskette mit RFID Tags ausgestattet werden, müssen sie billiger sein, weil sie nicht wiederverwendet werden können oder wollen. Wenn sich RFID Systeme noch mehr verbreiten, werden auch deren Kosten sinken und diese Technologie wird für immer mehr Anwendungen interessant werden.

### Ökonomische Faktoren

Durch die sich dauernd verschärfenden Wettbewerbsbedingungen und stärker werdenden Kostendruck ist Automatisierung ein wesentlicher Punkt durch den Unternehmen am Leben bleiben können. Auch die immer weiter wachsende Verknüpfung der Märkte und Branchen fördern die RFID Technologie. Die Transparenz der Supply Chain wird verbessert und die Transaktionskosten sinken.

#### Gesetzliche Vorschriften

Ein weiterer Punkt welcher die Technologie fördert, sind die Gesetze. Durch die immer globaler werdende Welt, die immer grösseren Anforderungen an Qualität und Dokumentation, verlangen die Gesetze immer bessere Rückverfolgbarkeit von Produkten. Im oben erwähnten Beispiel der Kennzeichnung von Nutztieren, können durch die RFID Technologie kranke Tiere genau zu den dazugehörenden Höfen zugeordnet werden und so Seuchenherde schnell lokalisiert und damit exakt eingegrenzt werden.
Sicherheit von Wartungsarbeiten (zum Beispiel Lüftungsklappen) ist Teilweise auch durch das Gesetzt geregelt. Ein betreibendes Unternehmen kann durch das Gesetz veranlasst werden, die verschiedenen durchgeführten Wartungsarbeiten genau zu dokumentieren. Damit kann ich Schadensfällen zum Beispiel Fahrlässigkeit oder menschliches Versagen evt. bewiesen werden. Eine solche Dokumentation dient somit auch zum Schutz des Betreibers, um sich vor ungerechtfertigten Schadensforderungen zu schützen.

### 3.4.2 Hemmende Faktoren

Natürlich gibt es bei sich neu etablierenden Technologien nicht nur positive Faktoren, welche helfen die Technologie zu verbreitet. Auch hemmende Faktoren sind bei dieser Technologie zu finden. Wie wir noch sehen werden, kann negatives Image viele Konsumenten verunsichern und die Technologie wird vielleicht vorschnell verteufelt.

#### Technische Probleme

Wie im Technologieteil schon erörtert wurde, führt die Nähe von Flüssigkeiten und Metallen zu Störungen bei der Datenübertragung. Dies stellt eine enorme Einschränkung der Verwendung von RFID dar. RFID Tags auf Getränkeverpackungen unterliegen Beispielsweise solchen Störungen. Diese Interferenzen und Frequenzverschiebungen treten jedoch nicht nur in der Nähe von Metallen und Flüssigkeiten auf, sondern können auch durch laufende Motoren entstehen. Ebenso können Abschattungseffekte durch fast gleichartige Signalfolgen auftreten. Der zweite Punkt wurde bereits in den fördernden Faktoren aufgelistet. Mit RFID ist es möglich Produktgruppen im Pulk zu scannen. Obwohl die Auslesezeit eines einzelnen Tags sehr schnell vor sich geht, kann es beim Auslesen von vielen Tags zu unakzeptablen Auslesezeiten kommen. Wenn ganze Paletten durch ein Auslesegerät gezogen werden können enorme Datenmengen anfallen, die in der Verarbeitung eine gewisse Zeit in Anspruch nehmen.

#### Standardisierung

Die Standardisierung im Bereich der RFID Technologie ist zwar fortschreitend, jedoch fehlen in einigen Bereichen die zwingend vorausgesetzten Standards um eine weltweite Etablierung erreichen zu können. Wie im technischen Teil bereits angesprochen existieren

heute die Standards von ISO und die des Auto-ID Centers. Global agierende Logistikunternehmen oder Handelsgruppen sind auf globale Standards angewiesen. Ein weiteres Problem das ebenso bei den technischen Problemen angesprochen werden könnte, sind die verwendeten Frequenzen. Sie müssen in allen Ländern freigegeben sein. Die UHF- und Mirkowellenbereiche in welchen die Transponder senden sind in einigen Ländern (noch) nicht für den kommerziellen Gebrauch freigegeben.

**Sicherheit der Datenübertragung**

Schon bei anderen Technologien (zum Beispiel Mobiltelefonen) hat man bei der Entwicklung dem Sicherheitsaspekt zu wenig Beachtung geschenkt, weil zu diesem Zeitpunkt natürlich keine Bedrohung für die Technologie vermutet wurde. Doch durch die alltägliche Verwendung einer Technologie können und werden sicherheitskritische Lücken auftauchen, welche bei der Einführung noch nicht bekannt waren. Eine Unternehmung welche eine neue Technologie adaptiert wird automatisch von ihr abhängig werden. Treten erst nach einer Zeit sicherheitskritische Lücken auf, werden dadurch wieder Investitionen anfallen, welche vielleicht von Anfang an vermeidbar gewesen wären. Es wird immer Individuen welche ihre Zeit dafür aufwenden, bei neunen Technologien die Schwachstellen zu finden und diese entweder auszunutzen oder sie zu veröffentlichen.

**Erfahrungsdefizite**

Noch besteht in den wenigsten Unternehmungen grosses Know how über die Verwendung von RFID. Das unternehmensspezifische Know how muss zuerst aufgebaut werden, was wiederum Investitionen mit sich bringt. Dieser Punkt bring grosse Unsicherheit in den betreffenden Unternehmen mit sich, was sich nicht gerade förderlich auswirkt.

**Datenschutz und Privatsphäre**

Datenschützer befürchten, dass mit der RFID Technologie der gläserne Kunde oder sogar der gläserne Bürger Wirklichkeit wird. So entstehen und kursieren zahlreiche Beispiele, wie Personen durch die Technologie ausspioniert werden könnten. Die grosse Unsicherheit der Konsumenten geht durch die Ubiquität der RFID Tags aus. In absehbarer Zeit werden die meisten Konsumgüter mit Tags ausgestattet sein [40]. Die RFID Technik ermöglicht den Missbrauch, da der Zugriff auf die Chips der Konsumgüter nicht wirksam beschränkt ist, und von jedem Interessenten unbemerkt mit einem Lesegerät ausgelesen werden kann. Da die EPC's der RFID einmalig sind können so die dazugehörigen Produkte eindeutig identifiziert werden. Werden nun die Produkte von Personen mitgeführt, sind auch sie eindeutig identifizierbar. Werden die Produkte mit einer Kreditkarte bezahlt wird aus einer identifizierbaren Person eine personalisierte Person. Auf diese Weise kann RFID in einfacher Weise dazu benutzt werden, sensible persönliche Daten über jeden einzelnen Verbraucher zu sammeln. Eine Möglichkeit um das Vertrauen der Konsumenten zu sichern, ist es, Geräte aufzustellen, mit denen es Konsumenten möglich sein wird, die Produkte zu anonymisieren, d.h. die Informationen auf den Tags zu löschen. Dazu muss jedoch

sichergestellt sein, das Produkte, welche mit einem RFID Tag versehen sind eindeutig gekennzeichnet sind, damit sich Konsumenten auch im klaren sind, das die Produkte überwacht werden können. Eine andere Möglichkeit, welche leider von RFID bis jetzt nicht unterstützt wird, wäre, dass die Konsumenten den Zugriff auf ihre RFID Tags beschränken könnten. Der Konsument wird also durch die Technologie eines seiner Rechte beraubt. Er ist nicht mehr in der Lage selber zu entscheiden in welchem Rahmen seine persönlichen Lebenssachverhalte offenbart werden.
Der Datenschutz wird jedoch von den Konsumenten als sehr wichtig betrachtet und kann so auch für den wirtschaftlichen Erfolg oder Misserfolg einer Technologie ausschlaggebend sein.

## 3.5 Zusammenfassung

Wie in dieser Seminararbeit aufgezeigt existieren viele Anwendungsgebiete für die RFID Technologie. Es konnte jedoch nur ein kleiner Teil der Anwendungsgebeite gezeigt werden. Viele weitere Anwendungen sind vorstellbar.

Auf der technologischen Seite sind viele Probleme gelöst. Es ist bereits möglich, einen günstigen RFID Tag zu produzieren und auf Produkten anzubringen, ohne das dessen Preis einen grossen Einfluss auf den Endpreis des Produktes hat. Was jedoch noch verbessert werden kann, ist zum Beispiel die Auslesegeschwindigkeit, die zum Teil Anwendungen noch nicht möglich macht. Ein weiteres Problem stellt die korrekte Erkennung der RFID Tags beim Auslesevorgang dar. Nach wie vor gibt es eine gewisse Fehlerkennung. Wenn diese Fehler bei tausenden billigen Produkten auftreten, kann man noch damit leben. Schwieriger wird es dann wenn zum Beispiel ein falscher Preis eines Produktes dem Kunden belastet würde. Ein anderer Verbesserungspunkt wäre ein Verschlüsselungsverfahren, um auch die Bedenken bezüglich dem Datenschutz zu zerstreuen.

Was die Kundenseite betrifft, so stehen sicher die Sicherheitsfragen bzw. -bedenken im Zentrum. Daher müssen mindestens folgende Anforderungen an die Verwendung der RFID Technologie aus Konsumentensicht bestehen [13].

- Der Käufer eines Produktes muss wissen, wenn sich darauf ein RFID Tag befindet. Er muss den RFID Tag entweder entfernen oder elektrisch zerstören können.
- Ohne das Wissen der betreffenden Person soll der RFID Tag nicht ausgelesen werden können. Dies ist ein Punkt der sehr schwierig zu erfüllen ist, beziehungsweise sogar unmöglich.
- Daten des Kaufes (zum Beispiel wenn der Kunde mit einer Kreditkarte bezahlt und dadurch eindeutig identifizierbar ist) dürfen nicht mit den Daten des RFID Tags kombiniert werden. So könnte ein eindeutiges Produkt eindeutig einem Käufer zugeordnet werden.

Auf der Produzentensicht bzw. der Anbietersicht werden die grössten Vorteile sicher in den Kosteneinsparungen gesehen. Die RFID Technologie wird sicher auch im Bereich der

Lebensmittelverwaltung immer eine grössere Rolle spielen und den Barcode mehr und mehr ablösen. Erst kürzlich hat sich die Migros und SAP zusammengetan und auf der Orbit-iEX den Einkaufsladen der Zukunft vorgestellt [20]. Es muss auch, wie hier gezeigt, noch viel auf der Ebene der Kundenbedenken bezüglich Privatsphäre diskutiert werden.

Wie in der Einleitung angetönt, ist der Mensch bezüglich neuen Technologien immer zurückhaltend. Es kann jedoch nur einen technologischen Fortschritt geben, wenn neue Technologien ausprobiert werden und die Menschen über ihren Schatten springen. Die Vergangenheit hat dieses Verhalten vielfach gezeigt und so wird auch die RFID Technologie ihr Dasein in der Zukunft stärken können.

# Literaturverzeichnis


[1] Universal Product Code (UPC) and EAN Article Numbering Code (EAN), Juni 2005
http://www.adams1.com/pub/russadam/upccode.html

[2] What is RFID, Juni 2005
http://www.rfidjournal.com/article/articleview/1339/1/129/

[3] AIM - The global trade association for automatic identification, Juni 2005
http://www.aimglobal.org/technologies/rfid/what_is_rfid.asp

[4] RFID Technology, Juni 2005
http://www.rfid-weblog.com/archives/RFID Presentation Sept 2004.ppt

[5] The History of RFID Technology, Juni 2005
http://www.rfidjournal.com/article/articleview/1338/1/129/

[6] Axcess Inc., Juni 2005
http://www.axcessinc.com/prod_radtag.php

[7] Tags get cheaper, Juni 2005
http://www.computerworld.com/softwaretopics/erp/story/0,10801,84004,00.html

[8] Information von Herr Prof. Stiller

[9] The Basics of RFID Technology, Juni 2005
http://www.rfidjournal.com/article/articleview/1337/1/129/

[10] A Summary of RFID Standards, Juni 2005
http://www.rfidjournal.com/article/articleview/1335/1/129/

[11] AIM - The global trade association for automatic identification, Juni 2005
http://www.aimglobal.org/standards

[12] CEN TC 278 Road Transport and Traffic Telematics, Juni 2005
http://www.aimglobal.org/standards/rfidstds/CENTC278.asp

[13] Controversy, Juni 2005
http://en.wikipedia.org/wiki/Rfid#Controversy

[14] Keeping a tab on livestock, Juni 2005
http://www.iso.org/iso/en/commcentre/pdf/Identification0101.pdf

[15] EPC-Netzwerk, Juni 2005
http://www.gs1-germany.de/internet/content/e39/e52/e140/e143/index_ger.html

[16] The compact Electronic Product Code, Juni 2005
http://www.autoidlabs.com/whitepapers/MIT-AUTOID-WH-008.pdf

[17] The EAN.UCC Code, Juni 2005
http://www.ean-int.org/get_intro.html

[18] Name service on the EPC network, Juni 2005
http://www.m-lab.ch/auto-id/SwissReWorkshop/papers/NameServiceOnTheEPCnetwork.pdf

[19] Physical Markup Language for Radio Frequency Identification, Juni 2005
http://xml.coverpages.org/pml-ons.html

[20] SAP und Migros lancieren RFID-Diskussion in der Schweiz, Juni 2005
http://www.inside-it.ch/frontend/insideit?XE7lhitk45Ym33WjzMZVZZzghtrLw4zN9
mc7zgdDDdheHlm8KhtLT8HQrIdg

[21] Tesco, Juni 2005
http://www.rfidjournal.com/article/articleview/658/1/1/

[22] Metro Group, Juni 2005
http://www.metrogroup.de

[23] Wal-Mart, Juni 2005
http://www.rfidgazette.org/walmart/

[24] Risiken und Chancen beim Einsatz von RFID, Juni 2005
http://www.bsi.bund.de/fachthem/rfid/studie.htm

[25] Shearwell Nutztier Identifikation, Juni 2005
http://www.shearwell.co.uk

[26] Chippflicht für Hunde, Juni 2005
http://www.admin.ch/cp/d/40d93942_1@fwsrvg.html

[27] Vorstellung des digitalisierten Reisepasses, Juni 2005
http://www.heise.de/newsticker/meldung/45780

[28] Neue Passgeneration in den Startlöchern, Juni 2005
http://www.heise.de/newsticker/meldung/47189

[29] Fussballweltmeisterschafts- Tickets, Juni 2005
http://www.heise.de/newsticker/meldung/43645

[30] Offizielle Webseite zur Fussballweltmeisterschaft 2006, Juni 2005
http://fifaworldcup.yahoo.com/06/de/

[31] RFID zur Wartung von Brandschutzklappen, Juni 2005
http://www.computerwoche.de/index.cfm?pageid=256&artid=52033&main_id=52033
&category=8&currpage=1&kw=

[32] Reifenhersteller Michelin, Juni 2005
http://www.rfidjournal.com/article/articleview/269/1/1/

[33] Nagra Homepage, Juni 2005
http://www.skidata.com

[34] Ressortlösung der Weissen Arena, Juni 2005
http://www.alpenarena.ch/pages/media/texte/07122001.asp

[35] Netzwelt.de Mauterhebung mittels RFID, Juni 2005
http://www.netzwelt.de/news/66876-erste-autonummernschilder-mitfidtechnologie.html

[36] Datasport: Zeitmessung bei Sportveranstaltungen, Juni 2005
http://www.datasport.com

[37] Diebstahlsicherung, Juni 2005
http://www.quarks.de/dyn/18314.phtml

[38] EuroID: Prozesssteuerung der Automobilindustrie, Juni 2005
http://www.euroid.com

[39] Metro Group: Supply Chain von Hühnereiern, Juni 2005
http://www.metrogroup.de/servlet/PB/menu/1014651_l1/index.html

[40] Datenschutz und RFID Technik, Juni 2005
http://ig.cs.tu-berlin.de/oldstatic/w2003/ir1/uebref/SuhrEtAl-Gutachten-G4-022004.pdf

# Chapter 4

# Fast Handover in Mobile IPv4

*Raoul Schmidiger*

*With the growing number of nomadic devices, Mobile IP came into born. Through different approaches and stragtegies, solutions were found to eliminate the problem of a Mobile Node to become offline during its movement through different subnets. These solutions, referred to as "'Fast Handover"', will be presented in the course of this seminary work.*

## Contents

## 4.1 Introduction

This seminary paper is an overview of Fast Handover mechanisms that can be used on existing network technology. It starts with a brief introduction on the IPv4 protocol. Then Mobile IP in general is introduced to the reader. Finally different methods to avoid a connection break of the Mobile Node to the Internet are presented.

## 4.2 Short IPv4 protocol description

IPv4 is the 4th version of the internet protocol (IP). It is the first version of the IP protocol, that was widely deployed and the Internet was developed on it. It is described in RFC 791 [1] and was published in 1981. It is designed for use in interconnected systems. It implements two basic functions: fragmentation and addressing.

It is used to transmit chunks of data that are called datagrams between source and destination. It provides mechanisms for fragmentation of large messages into datagrams that can be received asynchronously by the receiver and later be reasemmbled into the original message. This also allows different subnetworks to have different MTU's and the use of different source to destination routes for individual datagram packets. IP adresses have a fixed length out of four bytes. An address begins with a network part and finished with the local address that corresponds to an individual nodes physical network interface. This given, a fixed network topology results at any given point in time. The address space is therfore limited up to 4,294,967,296 unique addresses. Many of them are reserved for local networks and other things which further reduces the total real number of IP adresses that could be assigned to the public internet.

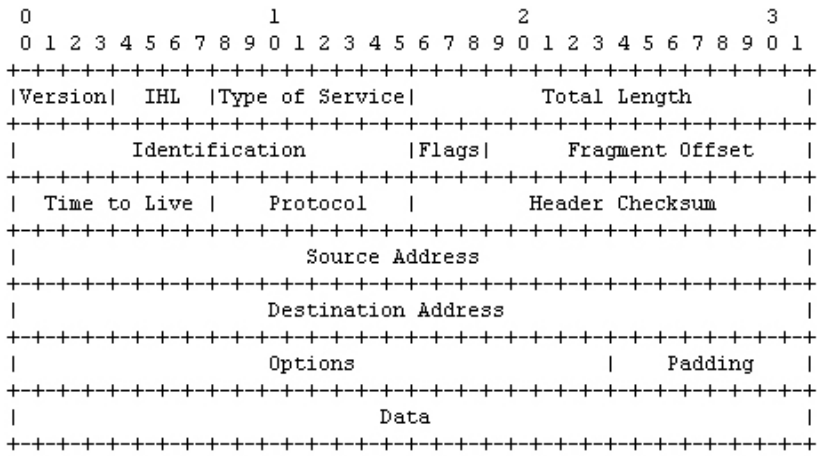


Figure 4.1: Example Internet Datagram Header

Short overview of the fields of an IPv4 Datagram Header:

- Version: Which version of the internet protocol is being used. 0100 for IPv4.

- IHL: Internet Header Length. Because the Options field does not have a fixed size, the actual length is stored here. Therefore an internet node can figure out, where the data starts.
- Type of Service: Not widely used today. May become more important in the future together with IPv6, when routing information should be stored in the datgram header as well. It could then be used to prefer certain routes to others, maybe less efficient routes.
- Total Length: The total length of the datagram packet, including header and data.
- Identification: Used for identification of the same IP packet.
- DF: Do not Fragment. If set, fragmentation is not allowed.
- MF: More Fragments to come. If set, more packets from the same IP packets will follow.
- Fragment Offset: Used to reassemble fragmented data. The offset where the data a fragment carries responds to the data in the original IP packet.
- Time to Live: A value that is decreased by one at every node the datagram passes by. If it reaches zero, the datagram is destroyed.
- Protocol: Specifies which type of protocol is used in the data part. Usually TCP or UDP.
- Header Checksum: Used for error recognition. Every router recalculates and sets this value, due to the change it does to the Time to Live value.
- Source address: The source IP address that issued this datagram as a 32 bit address.
- Destination address: The destination IP address of this datagram.
- Options: Used for special purposes like source-routing or timestamping. Has variable length.
- Padding: Since the Options field has variable length but the header always has to have a length of a mulitple of four, this fields is used to make up any resulting difference.
- Data: The actual payload.

## 4.3 Need for Handovers

A handover happens, when a mobile device changes in another network. The mobile device is then passed from the old network that supplied it with IP so far, to a new one. A nomadic device has to be reachable no matter in which subnetwork it is located at any given point in time. When the nomadic device moves it will eventually pass through

different subnetworks. At the network borders, a handover of the nomadic device from one subnetwork to the other has to happen, in order to keep the nomadic device connected to the overall network, the internet. This may not go without the user noticing it, but should happen automatically. Therefore the IETF supposed a protocol known as Mobile IP in IPv4 to realise this scenario.

### 4.3.1 Mobile IP in IPv4

In common IPv4 Networks, each node has a permanent physical attatchment point to the internet. A node in such a network is configured with either a static IP address or it obtains its IP address from a DHCP server. In both cases the prefix of the nodes IP address represents its network, its distinct place in the overall network topology.

However, Mobile IP, as specified by the IETF in RFC 3344 [2], requires for each mobile node to have a permanent IP address regardless of its current location in the various subnets and to allow seamless roaming between different IP subnetworks. Thus full transparency above the Link layer should be achieved. Furthermore, Mobile IP must not add any additional constraints on the assignement of internet addresses.

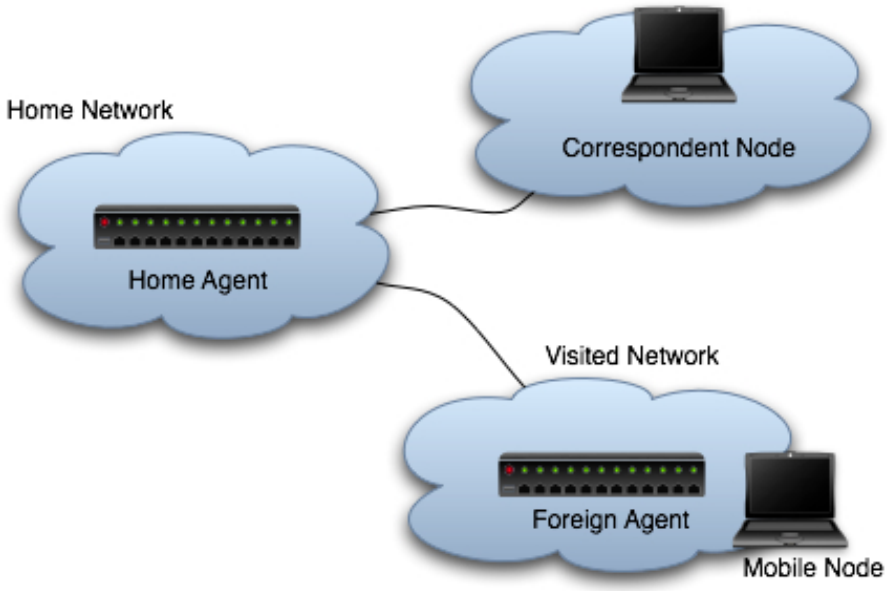


Figure 4.2: Basic Setup

In Mobile IP, each **Mobile Node (MN)** has a permanent IP address. This address is called the **Home address**. This Home address is usually set up by the owner of the mobile node or the administrator of his home network. Because network topologies are reflected in the IP address, it is not possible to keep a permanent IP address and connect to different subnets at the same time. Without Mobile IP, the user has to terminate all current sessions, release its current IP adress and obtain a new one either from a DHCP server or set it up manually. Then the user has to restart all sessions again. To solve

this issue for Mobile IP, **Home Agents** and **Foreign Agents (FA)** are utilized. Home Agents are located at the home network, which is the one used by the Mobile Node when considered "'at home"'. It is in fact the Home Agent that initially receives all requests that are destined to the Mobile Nodes permanent Home Adress. From there the requests are tunneled further on to the Foreign Agent, which acts as a router for the temproarily attatched Mobile Node.

In order to achieve this, the Home Agent has to keep track of the Mobile Nodes so called **Care-of Address (CoA)**. The Care-of Address is the current address of the Mobile Node in a foreign network. In order for this setup to work, the Mobile Node has to inform its Home Agent about its new Care-of Address, whenever it moves to a different network.

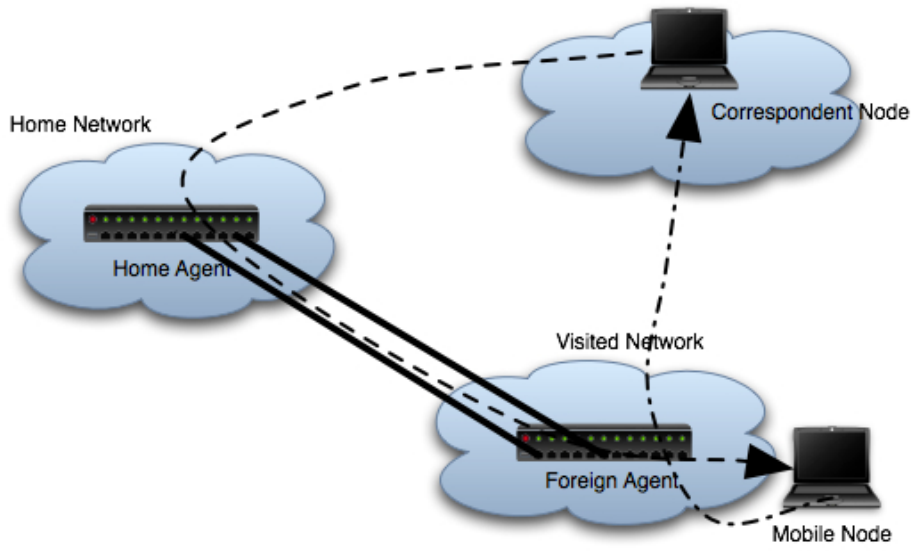


Figure 4.3: Triangular Routing

For the datagram to be sucessfully tunneled from the Home Agent to the Mobile Node, encapsulation of the original datagram is required. The original datagram has the Mobile Nodes permanent IP address in its destination header field, but from the Home Agent onwards, the current Care-of Address of the mobile node is required. This issue is solved by the Home Agent, which generates a new datagram with the current Care-of Address of the mobile node in its destination header field and the original datagram as its payload (see Fig.4.3). The Home Agent then tunnels the datagram to the Foreign Agent which detunels it and delivers it to the Mobile Node.

It the Mobile Node has to send packets back to the Corresponding Node, it uses the Corresponding Nodes IP, found in the encapsulated datagram, and uses the Foreign Agent as its router. This is called triangular routing.

However, triangular routing also has its downsides: routers and especially firewalls often only accept topologically correct adresses. But here te source address of the datagrams is the Mobile Nodes permanent IP address, although it should be the Care-of Address to be topologically correct. Another problem could be the Time to Live (TTL) field. The

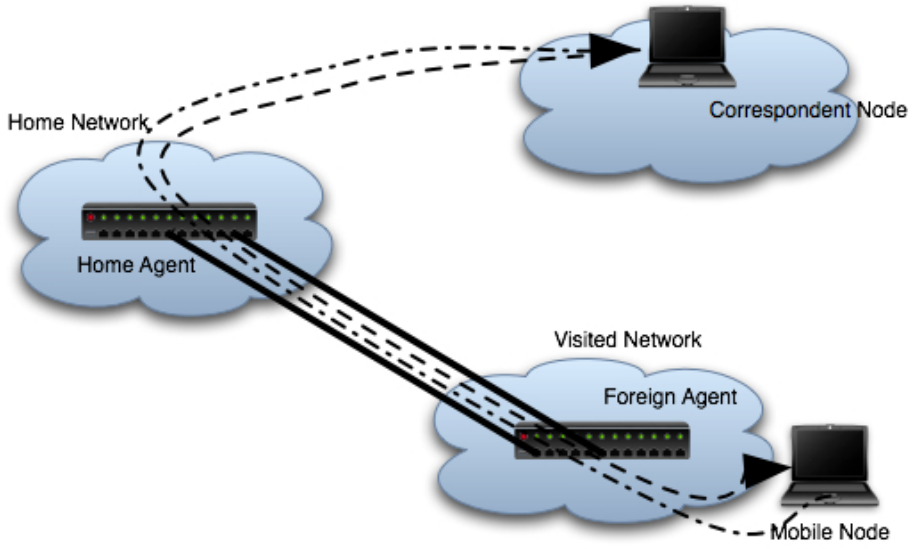


Figure 4.4: Reverse Tunneling

way to the mobile node may be shorter than the way from the mobile node and the TTL for the datagram not sufficient to reach the Corresponding Node.

Instead of Triangular Routing another approach can be taken: Reverse Tunneling. Here the way a Mobile Node received a datagram is simply reversed. In this case, the communication of the correspondent node and the mobie node remain unchanged. Now the datagrams are topologically correct and the TTL problem is also solved, since a tunnel always has the length of one. In this case the drawback lies in unoptimal, longer routes and hence higher network traffic.

### 4.3.2 Handover in IPv4

A Router that is configured to act as either a Home Agent or a Foreign Agent periodically sends so called Agent Advertisments through UDP packets. Agent Advertisments are an extension of standard ICMP Router Advertisement. When a Mobile Node receives such an Advertisment from a Foreign Agent, it sends a Registration Request back that includes the Home IP and the Home Agent IP. It is received by the Foreign Agent and relayed to the Home Agent which either grants or denies the request through a Registration Reply. Again the Foreign Agent processes the reply, adds some information to the registration and routes it on to the Mobile Node. In case the Mobile Node resides in its home domain already (co-located Care-of Address), the two intermediated steps, involving the Foreign Agent are skipped. On the other hand, if a Mobile Node does not receive any Agent Advertisment within some specified time threshold, it sends out Agent Solicitation messages itself. Either way, the goal is to register the Mobile Node to the Home Agent in the end.

This process has to happen whenever the mobile node moves to a different network, and after the Data Link Layer (Layer 2, OSI Model) has been established to the new Foreign

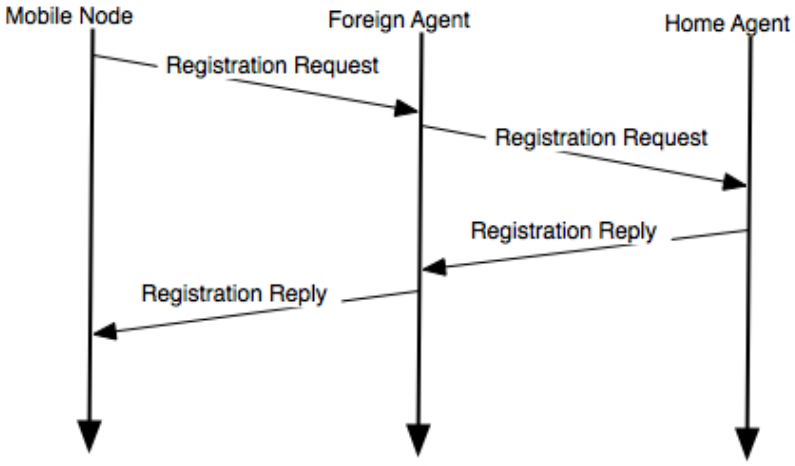


Figure 4.5: Handover in foreign Network

Agent. Strictly speaking, this is not an active handover from one Foreign Agent to the other but rather just a restart of the usual binding procedure.

## 4.4 Problems that can arise when changing a network

When the nomadic device travels from one subnetwork to another, the handover happens once the data link layer to a new Foreign Agent is established (Layer 2 handover). The registration does not take zero-time though, as the registration request has to propagate through the network.

Before the binding is completed, the Mobile Node is not able to either send or receive datagrams. This may be not be a problem, when using a browser to surf the internet, requesting a new page only every now and then. However, when connected to eg. a video stream, the stream will brake, once the Data Link Layer switches. Any application relying on a steady transport layer (TCP, UDP, SCTP, DCCP, ...) will notice the handover and act in an unpredictable fashion.

To avoid a connection loss, Fast Handover is be used.

### 4.4.1 Fast Handover

The problems that normal handover introduce, are solved through two distinct methods known as Pre- and Post-Registration Handover. These are two methods to achieve what is called Fast Handover. Fast Handover means, that the Mobile Nodes ability to send and to receive packets, is never absent during Handover.

The Pre-Registration method allows a Mobile Node to perform a Network Layer (Layer 3) Handover, before the Data Link Layer (Layer 2) handover is done. When the Data Link

Layer is switched to the new Foreign Agent, no more Registration overhead is required, the Home Agent has already binded the Mobile Node to the new Care-of Address at the new Foreign Agent.

The Post-Registration method allows the Mobile Node to receive data through the new Foreign Agent, even before the formal Registration to the Home Agent is completed or even initiated.

These two methods can also be used in combination. They are invoked through the triggers on the Data Link Layer. These triggers are received by either the Mobile Node (Mobile Trigger), received by the old Foreign Agent (Source Trigger) or received by the new Foreign Agent (Target Trigger). They contain information on the Foreign Agents and the Mobile Nodes IP address. The following explanations are based on Yan Wei's paper on Fast Handover in IPv4 [3] to a high degree.

### Pre-Registration

The Pre-Registration method can be triggered by all three "'types"' of trigger: Mobile, Source, and Target Triggered. For performing a successful Fast Handover it is required that there is enough time left for the Layer 3 handover to complete after the trigger is received and the actual Layer 2 handover takes place. Additionaly, the latency can be reduced through caching mechanisms in Foreign Agents in such a way, that each Foreign Agent periodically sends out Router Solicitation Requests to its neibouring Foreign Agents and stores the last received Router Advertisments Replies. When a Mobile or Source Triggered network change is initiated, that step is already done.

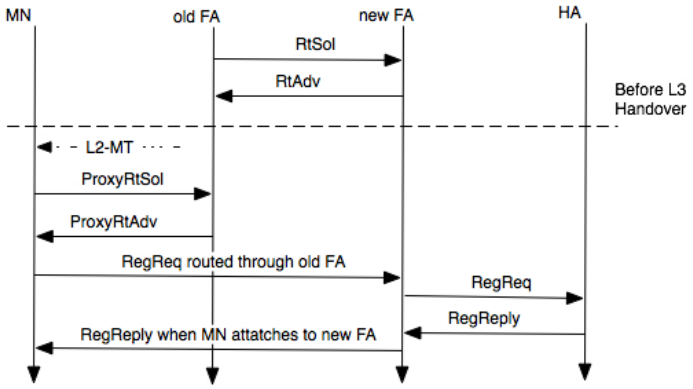


Figure 4.6: Mobile Triggered Handover

**Mobile Triggered** The Mobile Node receives a Trigger, informing it, that it will soon move to a new Foreign Agent. It then sends a Proxy Router Solicitation to the old Foreign Agent. By then the Foreign Agent will usually already have sent a Router Solicitaition to the new Foreign Agent and passes the new Foreign Agents Router Advertisment answer, as a Proxy Router Advertisment to the Mobile Node. Then the Mobile Node sends a Registration Request through the old and the new Foreign Agents to the Home Agent.

The Home Agent sends the Registration Reply back to the new Foreign Agent and from there to the Mobile Node, once Layer 2 handover is completed.

**Source Triggered** The old Foreign Agent receives a Source Trigger when the Mobile Node is going to move to another Foreign Agent. If the old Foreign Agent has not already sent a Router Solicitation to the new Foreign Agent, it will do so now, adding more latency to the Layer 3 handover. Usually that has already happend, before the Trigger is received by the old Foreign Agent. The Routing Advertisment sent back from the new Foreign Agent to the old Foreign Agent is forwareded to the Mobile node as a so called Proxy Routing Advertisment. The Mobile Node can now send a Registration Request to the Home Agent through the old Foreign Agent, that routes it through to the new Foreign Agent. This way is taken because the Mobile Node is not yet connected to the new Foreign Agent, as no Layer 2 handover has happend yet. The Home Agent sends the Registration Reply back the same way down to the new Foreign Agent. After Layer 2 handover is finished, the Registration Reply is delivered to the Mobile Node.

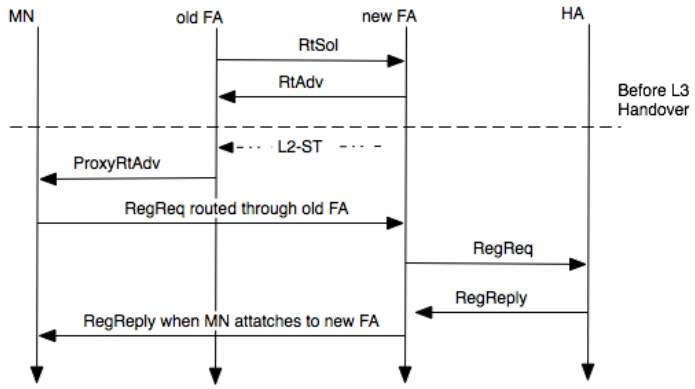


Figure 4.7: Source Triggered Handover

**Target Triggered** The new Foreign Agent receives a Trigger form the old Foreign Agent. The new Foreign Agent knows a Mobile Node is soon to change into its network and treats the trigger as a Router Solicitation and tunnels its Agent Advertisment through the old Foreign Agent to the Mobile Node. The Mobile Node then sends a Registration Request through the old Foreign Agent to the new Foreign Agent and from there to the Home Agent. The Home Agent sends the Registraton Reply back to the new Foreign Agent and from there to the Mobile Node, once the Layer 2 handover has been completed.

### Post-Registration

The Post-Registration method allows a Mobile Node to send and receive datagrams before Registration to its Home Agent through the use of unidirectional or bidirectional tunnels between the old and the new Foreign Agent. Because the Mobile Node does not require to bind to its Home Agent at all, while on a new Foreign Agents subnetwork, it is here also theoretically possible, that a Mobile Node moves not only to one new Foreign Agent, but

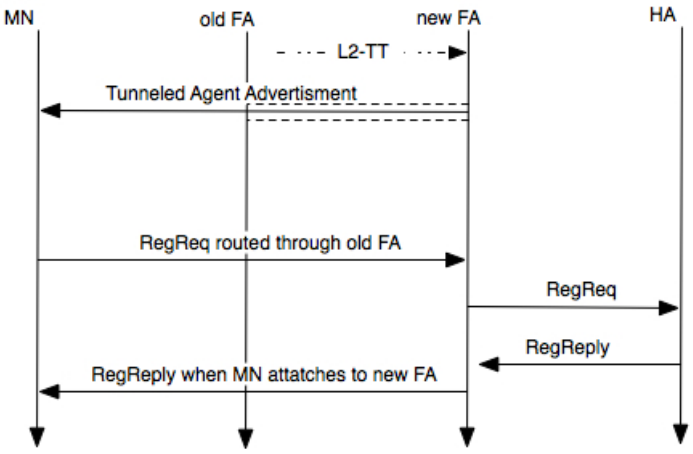


Figure 4.8: Target Triggered Handover

also to a second and so on. Therefore, the old Foreign Agent is refered to as the anchor Foreign Agent when more than two Foreign Agents are involved. Therefore two methods have to be considered when talking about Post-Registration: two-party and three-party Post-Registration Handover. For each of theses methods, Source and Target Triggered scenarios exist.

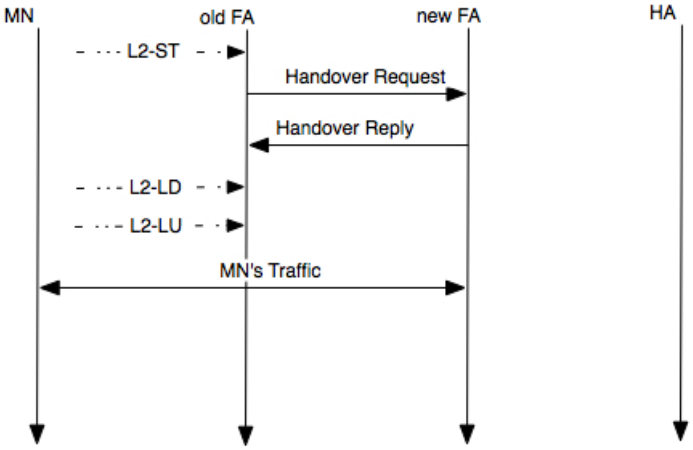


Figure 4.9: Source Triggered Two Party Handover

**Source Triggered Two-Party Handover** The old Foreign Agent receives a trigger that a Mobile Node is about to move from to a new Foreign Agent. The old Foreign Agent then sends a Handover Request to the new Foreign Agent, which is identified from the information contained in the trigger. The new Foreign Agent sends a Handover Reply. If in these to messages both agreed to set up a tunnel, the Layer 2 handover is initiated. When the Layer 2 handover is completed, the old Foreign Agent tunnels packets destined to the Mobile Node to the new Foreign Agent and from there to the Mobile Node. Packtes from the Mobile Node are delivered to the new Foreign Agent and then are either tunneled backwards to the old Foreign Agent or they are sent on using

normal routing (unidirectional tunnel). If the Mobile Node then issues a Registration Request successfully, the new Foreign Agent becomes the anchor Foreign Agent. Three-party handover happens, when two-party handover has already happend, but before the Mobile Node registered to its Home Agent and the Mobile Node moves to a third Foreign Agent.

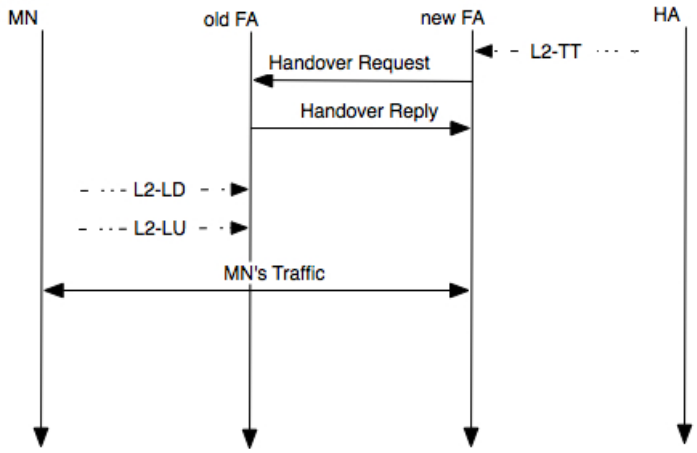


Figure 4.10: Target Triggered Two Party Handover

**Target Triggered Two-Party Handover** The new Foreign Agent receives a trigger, informing it, that a Mobile Node is moving from an old Foreign Agent its network. This time, the new Foreign Agent sends a Handover Request to the old Foreign Agent. If both Foreign Agents agree to set up a tunnel, the procedure is the same as in Source Triggered two-party Handover.

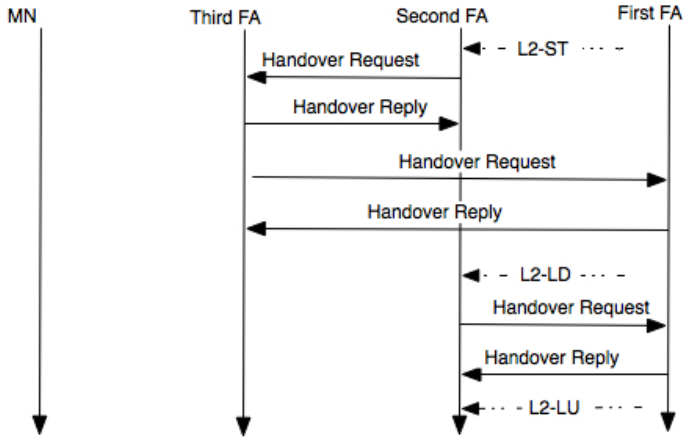


Figure 4.11: Source Triggered Three Party Handover

**Source Triggered Three-Party Handover** The new Foreign Agent receives a trigger, that the Mobile Node will move to a third Foreign Agent and issues a Handover Request to that Foreign Agent. The third Foreign Agent then sends a Handover Request to the

old or anchor Foreign Agent to initiate tunnel. After the new tunnel is set up, the second or new Foreign Agent issues a Handover Request to the old or anchor Foreign Agent to release the tunnel. After establishment of the tunnel between the third Foreign Agent and the anchor Foreign Agent, the Layer 2 handover is initiatied.

**Target Triggered Three-Party Handover** The thrid Foreign Agent receives a trigger, that a Mobile Node will enter its subnetwork and issues a Handover Request to the new or second Foreign Agent. Then the same happens as in Source Triggered handover. In both cases the intermediate Foreign Agent is removed and the old or anchor Foreign Agent remains the one end of the tunnel communicating with the Home Agent, as long as the Mobile Node did not yet complete a successful new Registration Request while travelling through the various different subnetworks.

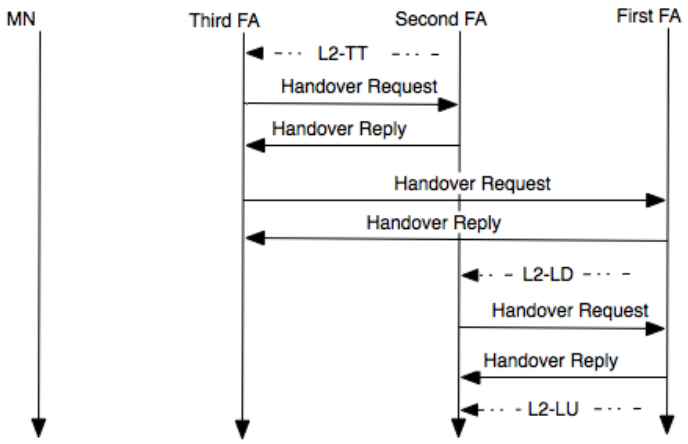


Figure 4.12: Target Triggered Three Party Handover

#### Combined Registration Method

Here Pre- and Post-Registration techniques are used at the same time.

If the Layer 3 handover completes before the Layer 2 handover, that was it. Otherwise Post-Registration method makes sure, the Mobile Nodes connection is maintained. Although the Combined Registration Method has some overhead in case Pre-Registration works, it is certainly a neat solution.

## 4.5 Summary

Before Mobile IP was introduced, a lot of hardship had to be taken into account when moving with a nomadic device through different subnetworks. Configuration of the device was necessary whenever a new subnet should be used to connect to the internet. With Mobile IP and correspondingly configured routers, this issue was solved. However only

Fast Handover mechanisms made the transition from one subnetwork to another completly transparent, not only to the user, but also to all layers above the Network Layer. The principles are simple and straight-foreward. However, Fast Handover is completly redundant, when a Mobile Node doesn't have a wireless connection or when wireless networks don't overlap... And yes, that is the case in probably 95% of the earths surface and will unlikely ever change.

# Bibliography

[1] RCF 791:Internet Protocol - DARPA Interntet Program Protocol Specification

[2] RFC 3344:IP Mobility Support for IPv4

[3] Yan Wei: Fast Handover Mechanisms in Mobile IPv4 Networks, http://www.tml.hut.fi/Studies/T-110.551/2003/papers/10.pdf, May 2003.

# Kapitel 5

# The Session Initiation Protocol in Mobile Environments

*Ruben Meier und Lorenz Fischer*

*In diesem Kapitel geht es um SIP (Session Initiation Protocol), SDP (Session Description Protocol) und RTP (Real-Time Transport Protocol). Alles Protokolle welche vor allem in der Voice-over-IP (VoIP) Telefonie verwendet werden. SIP und SDP sind Protokolle für den Verbindungsaufbau multimedialer Konferenzen zwischen zwei oder mehreren Teilnehmern. SIP wurde von der Internet Engineering Task Force (IETF) speziell für das Internet entwickelt und ist bekannten Protokollen wie HTTP oder SMTP sehr ähnlich. RDP ist ein Protokoll, welches für die Übertragung von multimedialen Daten verantwortlich ist. Voice-over-IP und somit auch SIP erleben zur Zeit einen starken Aufschwung. Wir zeigen, welche Funktionen SIP verwendet und welche Soft- und Hardwarekomponenten diese beherrschen. Weiter werden wir die wichtigsten Vor- und Nachteile dieser relativ neuen Technologie diskutieren. Da SIP jedoch nicht nur für VoIP-Systeme genutzt werden kann, werden auch allgemeine Aspekte von SIP im Bezug auf Anwendungsgebiete, Sicherheit und Mobilität zur Sprache kommen. Neben den technischen Aspekten versuchen wir jedoch auch die möglichen Veränderungen im Kommunikationsverhalten durch auf SIP basierenden Systemen aufzuzeigen. So werden anhand verschiedener Beispiele Systeme welche bereits erfolgreich im Einsatz stehen, aber auch Visionen und möglichen Weiterentwicklungen der Zukunft vorgestellt.*

## Inhaltsverzeichnis

## 5.1 Einleitung

### 5.1.1 Was ist SIP?

Das Session Initiation Protocol (SIP) ist ein von der *Internet Engineering Task Force (IETF)* entwickelter Standart für multimediale Konferenzen über das IP Protokoll. SIP ist ein auf der Anwendungsschicht arbeitendes Kontrollprotokoll. Ähnlich wie HTTP oder SMTP arbeitet SIP mit dem normalen ASCII Zeichensatz. Auch die verwendete Headerinformation ähnelt grösstenteils derjenigen dieser zwei Protokollen. Mit SIP lassen sich Anrufverbindungen erstellen, unterhalten und beenden wobei fünf verschiedene Teilaspekte unterschieden werden:

- Standort des Benutzers: Bestimmen welche Systeme für die Kommunikation zur Verfügung stehen.
- Verfügbarkeit der Teilnehmer: Ermitteln ob die angerufenen Personen bereit sind an einer Kommunikation Teilzunehmen.
- Technische Möglichkeiten der Benutzer: Ermitteln welche Medientypen und Medienparameter verwendet werden können.
- Verbindungsaufbau: Klingeln und das Einrichten der Sitzungsparameter auf beiden (allen) Seiten.
- Verbindungsverwaltung: Transfer und Beendigung von Sitzungen, Verändern von Sitzungsparametern oder das Aufrufen von Diensten

Die *Session Initiation Protocol working group* arbeitet weiterhin daran das SIP Protokoll zu erweitern und versucht wenn immer möglich auf Vorschläge Dritter einzugehen. Sie betreibt jedoch nicht aktiv Forschung für die Anwendung von SIP in spezifischen Hard- und Softwareumgebungen. Weiterentwicklungen sollen sich an eine von der Entwicklergruppe vefassten Charta halten:

1. Dienste und Funktionen werden direkt zwischen zwei Endpunkten betrieben. Ausser für den Verbindungsaufbau sollten also möglichst wenige Systeme in den Kommunikationsprozess involviert sein. Dies verringert die Netzwerkauslastung.
2. Erweiterungen und Dienste müssen generell anwendbar sein und verschiedene Arten von Kommunikationssitzungen unterstützen. Systembestandteile, welche zum Beispiel nur für Videokonferenzen verwendet werden können, gilt es zu vermeiden.
3. Einfachheit als oberstes Ziel: je einfacher ein Protokoll spezifiziert wird, umso geringer ist die Anzahl Fehlerquellen bei der Implementation und umso grösser ist die Chance der Akzeptanz.
4. Wiederverwendbarkeit bestehender IP Protokolle und Architekturen sowie die Integration in die bestehenden Datennetze wird angestrebt.

Die komplette Charta steht im Internet unter zum Download bereit. Der gesamte Funktionsumfang von SIP wird in RFC 3261 [2] spezifiziert.

### 5.1.2 SIP Lebenslauf

- 1996: Henning Schulzrinne Professor an der Columbia University und seine Mitarbeiter machen erste Vorschläge welche an die IETF geschickt werden. Das darin beschriebene *Simple Conference Invitation Protocol (SCIP)* hat zum Ziel, Benutzern eine standardisierte Möglichkeit zu bieten andere Personen zu Multimediakonferenzen einzuladen.
- 1996: die IETF erlässt die erste SIP Spezifikation in RFC 2543 [3].
- 2000: SIP wird im November offiziell als Signalisierungsprotokoll für UMTS ausgewählt.
- 2002: Weiterentwicklung des Standards und Erlass einer weiteren SIP Spezifikation in RFC 3261 [2] welche erstere ersetzt. Mit der Einführung dieser Spezifikation wurde der Glaube an eine Fortführung der Entwicklungsarbeiten an SIP gestärkt. Viele Arbeiten für verschiedene Bereiche der SIP-Kommunikation wurden erstellt und in weiteren RFCs spezifiziert. Als Beispiele seien hier RFC 3262, welches sich mit der Ausfallsicherheit von provisorischen Antworten beschäftigt, oder RFC 3263, welches Regeln für das Auffinden von SIP Proxyservern beschreibt, erwähnt.

## 5.2 Funktionen/Operationen

### 5.2.1 Überblick

*SIP* hat die Aufgabe eine Verbindung zwischen verschiedenen Kommunikations-Partnern aufzubauen. *SDP* wird dazu verwendet, die während der Kommunikation übertragenen Daten zu beschreiben. Per *RTP* werden dann die eigentlichen Daten übermittelt. Alle diese Protokolle werden im folgenden vorgestellt und erklärt.

Grundsätzlich sind folgende Schritte in einem Verbindungs-Protokoll notwendig:

Tabelle 5.1: notwendige Schritte für ein Verbindungs-Protokoll

| **Funktion** | **SIP-Funktion** | **SDP-Funktion** |
|---|---|---|
| Registrierung | Register<br>Client-Registrar | -<br>- |
| Verbindungsaufbau | Invite<br>Clients, Proxys, Registrar | Medien- Informationen<br>Clients |
| Verbindung | Re-Invite<br>Clients, Proxys, | Medien- Informationen<br>Clients |
| Verbindungsabbau | Bye<br>Clients, Proxys | - |

Um eine Verbindung zwischen zwei Partnern zu erstellen, müssen sich beide Teilnehmer bei einem Provider registrieren lassen. Dazu hat sich jeder bei einem Kontroll-Knoten,

einem so genannten Registrar, mit der Register-Methode zu registrieren. Dadurch können sie in Zukunft über eine eindeutige Adresse lokalisiert werden. Diese eindeutigen Adressen werden URI (Uniform Ressource Identifier) genannt.

### 5.2.2 Das SIP-Protokoll (RFC-3261)

In der RFC 3261 [2] sind die wichtigsten SIP Methoden erklärt. Der Verbindungsaufbau läuft immer über einen Proxy. Alle Methoden werden von diesem Proxy an die nächste Stelle weitergeleitet. Genaueres über die Funktion eines Proxys folgt unter 5.3.2.

#### SIP URI

Eine SIP URI (Uniform Resource Identifier) ist eine für SIP verwendete Adresse eines Benutzers. Die Form ähnelt der einer Emailadresse. Sie besteht aus einem Benutzernamen und einem Computernamen, welche durch ein @ getrennt werden. Um die Adresse von einer Emailadresse unterscheiden zu können, wird ihr der String *sip:* vorangestellt. Eine gültige SIP-Adresse wäre also zum Beispiel: *sip:alice@atlanta.com*. Um einen Benutzer über eine verschlüsselte Leitung anzurufen, verwendet man anstelle der normalen URI eine sichere URI, SIPS URI genannt. Ein Beispiel dafür wäre *sips:alice@atlanta.com*. Mehr über sichere Verbindungen mit SIP folgt in 5.6. Ein mit einer SIP URI identifizierbarer Benutzer wird *Useragent* oder kurz *UA* genannt.

Sobald sich ein Benutzer an einem SIP-Netzwerk registriert, generiert sein System eine *umgebungsabhängige SIP-Adresse* und verwendet dazu den Computernamen der Maschine, an der er zur Zeit arbeitet. Da nicht jeder Computer über einen eingetragenen Domänennamen verfügt, kann eine SIP-Adresse auch die Form *Benutzer@IP-Adresse* haben (z.B. bob@192.168.0.20). Wenn sich ein Benutzer jedoch bei einem SIP-Anbieter anmeldet, wird ihm eine SIP-Adresse in der Domäne des Anbieters zugewiesen. Diese Adressen werden dann *umgebungsunabhängig* genannt, weil der Benutzer, egal von wo er sich ins Internet einwählt, unter *Benutzer@Anbieterdomäne* erreichbar sein wird.

#### INVITE

Die invite Methode sendet eine Einladung an einen anderen SIP- Benutzer. Die Invite Methode wird von allen benötigten SIP-Proxies weitergeleitet und landet am Ende beim Ziel, ausser der Verbindungsaufbau wird abgebrochen. Der Header einer typischen Invite-Nachricht sieht folgendermassen aus [2]:

```
INVITE sip:bob@biloxi.com SIP/2.0
Via: SIP/2.0/UDP pc33.atlanta.com;branch=z9hG4bK776asdhds
Max-Forwards: 70
To: Bob <sip:bob@biloxi.com>
From: Alice <sip:alice@atlanta.com>;tag=1928301774
Call-ID: a84b4c76e66710@pc33.atlanta.com
```

```
CSeq: 314159 INVITE
Contact: <sip:alice@pc33.atlanta.com>
Content-Type: application/sdp
Content-Length: 142
(Die SDP-Abschnitte fehlen)
```

Während die meisten dieser Headerzeilen einigermassen selbsterklärend sind, haben andere ungewohnte oder neue Funktionen. Die Zeile *Via: ..* enthält gleich mehrere Informationen. Einerseits die Adresse, an der Alice die Antwort auf diese Invitenachricht erhalten möchte, und andererseits Protokollinformationen. Am Schluss der Zeile wird der Parameter *branch* angegeben der diese INVITE-Transaktion identifiziert. Jeder Proxy den diese Nachricht passiert, fügt eine eigene Via-Zeile in den Header der Nachricht. So kann der *Heimweg* für darauffolgende Nachrichten einfacher gefunden werden. Das Headerfeld *Call-ID* beinhaltet einen global eindeutigen Bezeichner für diese Verbindung. Die Kombination aus From-Adresse, To-Adresse und Call-ID identifiziert die mit SIP gestartete Verbindung eindeutig. *CSeq* besteht aus einer Ganzzahl gefolgt von einem Methodennamen. Die Zahl wird bei jedem Befehl um eins erhöht und ist eine normale Sequenznummer. Das Feld *Contact* enthält die direkte Adresse zur Maschine von Alice. Bei dieser Adresse sind IP-Adressen anstelle von Hostnamen erlaubt.

### re-INVITE

Eine während einer bestehenden Verbindung gesendete INVITE-Nachricht wird als Re-Invite bezeichnet. Durch diese können die Eigenschaften einer bestehenden Verbindung geändert werden. Sie referenziert die bestehende Verbindung über die bereits bestehende CALL-ID. Dadurch könnte beispielsweise aus einem anfänglichen Telefonat eine Videokonferenz gemacht werden. Die re-INVITE-Nachricht wird entweder mit einem OK oder einem Fehlercode beantwortet. In jedem Fall wird diese Antwort jedoch mit einem ACK vom Initianten quittiert. Darauf werden die neuen Einstellungen entweder übernommen oder die Kommunikation wird mit der bereits ausgehandelten Konfiguration fortgesetzt.

### REGISTER

Benutzer melden sich beim Registrar per REGISTER-Methode an. Dieser Vorgang wird periodisch wiederholt um die Daten des Registrars aktuell zu halten. Beispielhaft ist die REGISTER-Methode in Abbildung 5.1 dargestellt: Es zeigt ein Szenario in welchem sich Benutzer Alice beim Registrar *iptel.org* registriert. Die übermittelte Adresse von Alice wird auf einem *Location-Service-Server* gespeichert. Registrar und Location-Service können, müssen aber nicht, auf zwei physisch unterschiedlichen Maschinen arbeiten. Die Proxyserver fragen beim Suchen nach SIP-UA's jeweils bei den zuständigen Location-Service-Diensten nach der aktuellen IP der gesuchten Person. Dies kann zum Beispiel über das LDAP Protokoll geschehen. Die Schritte 1 und 2 des Bildes 5.1 genügen vollkommen für die REGISTER Methode. Die anderen Schritte zeigen, wie Benutzer Bob Alice über seinen Proxy finden kann.

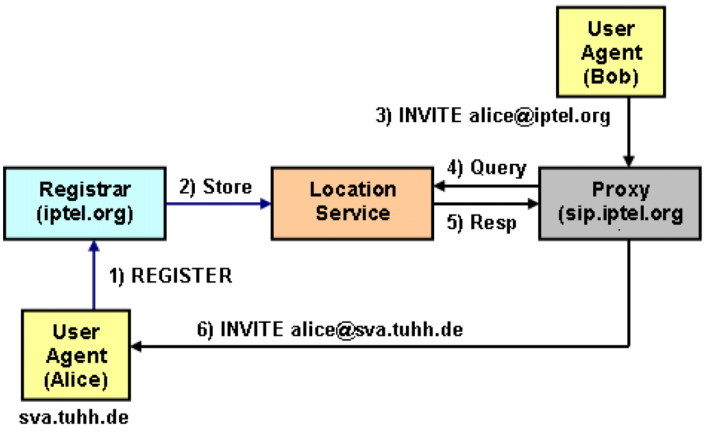


Abbildung 5.1: REGISTER [21]

### ACK

SIP benutzt einen Drei-Weg-Handshake: Nach dem Senden der INVITE-Nachricht wird diese mit einem OK bestätigt. Den Eingang der OK-Nachricht wird durch ein ACK quittiert. Der Befehl ACK ist immer eine Bestätigung auf eine Antwort und es bestehen deshalb bestimmte Abhängigkeiten bezüglich der Daten die eine ACK-Nachricht im Header tragen muss: Das Kopffeld muss in den Feldern Call-ID, From, und Request-URI die gleichen Werte beinhalten wie der Header der **Anfrage-Nachricht**. Das *To*-Feld der ACK-Nachricht muss aus dem gleichen Wert bestehen, wie das *To*-Feld der **Antwort**, die mit dem ACK bestätigt wird. Weiter darf die Bestätigung nur eine einzige *Via*-Zeile beinhalten und diese muss mit der Zeile der ursprünglichen Anfrage identisch sein. Zuletzt muss sichergestellt werden, dass die *CSeq-Nummer* mit der Zahl in der Anfrage-Nachricht übereinstimmt. Als Methodenname muss jedoch *ACK* in der CSeq-Zeile stehen.

### Cancel

Cancel bewirkt den Abbruch einer beliebigen Transaktion. Jedes Gerät (ausser den Stateless-Proxies) einer SIP-Umgebung wird auch als *Transaction User* oder kurz *TU* bezeichnet. Wenn ein TU eine Anfrage an ein anderes Gerät starten will, generiert es zunächst eine Transaktion. Diese Transaktionen können durch den CANCEL-Befehl abgebrochen werden. Wenn zum Beispiel bei der Suche nach einem SIP-Benutzer zu viele Um- und Weiterleitungen gemacht werden kann es sein, dass die maximale Anzahl Weiterleitungen überschritten werden. Diese Anzahl wird durch das Max-Forwards-Feld im Header der SIP-Nachricht definiert. Beim Erreichen der maximalen Menge von Weiterleitungen, wird die Anfrage und somit die Transaktion per CANCEL abgebrochen. Die generierte CANCEL-Nachricht wird dann über alle bereits passierten Proxies zurückpropagiert.

### BYE

Bye wird übermittelt, um eine bestehende Verbindung zu beenden. Generiert wird die BYE-Nachricht wie jede andere Nachricht im SIP-Protokoll. Wenn eine Seite die laufende Verbindung abbrechen möchte, schickt Sie dem Gegenüber ein BYE. Ab dem Versand der BYE Nachricht darf der Sender keine weiteren Datenpakete annehmen, sondern muss auf eine Antwort warten. Der Erhalt einer BYE-Nachricht muss mit einem OK bestätigt werden. Auf das OK wird kein erneutes ACK versendet. Wenn keine Antwort auf eine gesendete BYE-Nachricht eingeht, kann der Sender von einem Timeout ausgehen und die Verbindung beenden.

### OPTIONS

Ein Benutzer kann durch die OPTIONS-Methode die Möglichkeiten eines anderen Benutzers abfragen. Auch die OPTIONS-Nachricht wird wie jede andere Nachricht generiert. Im Header der Anfrage kann ein *Contact*-Feld stehen, um bekannt zu geben, wer sich für die angefragte Information interessiert. Weiter sollte der Header ein *Accept*-Feld besitzen, um der Gegenstelle bekanntzumachen, welche Informationen man selbst verarbeiten kann (z.B. SDP). Ein Beispiel einer OPTIONS-Nachricht sieht wie folgt aus [2]:

```
OPTIONS sip:carol@chicago.com SIP/2.0
Via: SIP/2.0/UDP pc33.atlanta.com;branch=z9hG4bKhjhs8ass877
Max-Forwards: 70
To: <sip:carol@chicago.com>
From: Alice <sip:alice@atlanta.com>;tag=1928301774
Call-ID: a84b4c76e66710
CSeq: 63104 OPTIONS
Contact: <sip:alice@pc33.atlanta.com>
Accept: application/sdp
Content-Length: 0
```

### SIP-Codes

SIP übermittelt ähnliche Status-Codes wie das HTTP-Protokoll. Es gibt zu viele Status Codes, um alle aufzuzeigen, sie sind alle in RFC 3261 [2] beschrieben. Die Codes werden in die folgenden Hauptgruppen unterteilt:

- 1xx: Provisional – Die Anfrage wurde empfangen und ist in Verarbeitung.
- 2xx: Success – Die Anfrage wurde komplett empfangen und wurde verarbeitet.
- 3xx: Redirection – Weitere Schritte müssen ausgeführt werden, um die Anfrage zu erfüllen.
- 4xx: Client Error – Der Server kann eine fehlerhafte Anfrage vom Client nicht verarbeiten.

- 5xx: Server Error – Der Server kann eine wahrscheinlich korrekte Anfrage vom Client nicht verarbeiten.
- 6xx: Global Failure – Die Anfrage kann bei keinem Server erfüllt werden.

### 5.2.3 Das SDP-Protokoll (RFC-2327)

Während SIP für die Verbindungsverwaltung zuständig ist, beschreibt SDP das Format der zu übertragenden Nutzdaten. Die wichtigsten Methoden sind in RFC 2327 [4] zu finden. Folgende Session-Eigenschaften werden von SDP beschrieben:

- Name und Zweck der Sitzung
- Zeit, während der die Sitzung aktiv ist
- Die Art der Medien, die in der Sitzung verwendet werden
- Detailinformation über diese Medien (Adressen, Ports, Formate, ...)
- benötigte Bandbreite
- Kontaktperson dieser Sitzung

Bei jedem SIP-Verbindungsaufbau wird ein SDP-Paket mitgeliefert um das gewünschte Medium zu bezeichnen. Nur SIP ohne SDP würde nicht ausreichen, um eine Multimedia-Übertragung zu starten. Generell muss in der SDP-Nachricht soviel Information enthalten sein, wie nötig ist, um einer Sitzung beizutreten. Ausgenommen davon sind Sicherheitsschlüssel. Die wichtigsten Aspekte werden im Folgenden noch genauer beschrieben.

#### Medieninformation

Die Detailinformation einer SDP-Nachricht muss folgende Angaben beinhalten:

- Der Typ des Mediums (Video, Audio, Chat oder anderes)
- Das Format des Mediums (verwendete Kompression und Codecs inklusive derer Versionsnummern)
- Das verwendete Transportprotokoll (RTP, UDP, IP, H.320 oder andere)
- Der verwendete Port auf dem gesendet wird (nur für IP basierte Netze)
- Weitere für das verwendete Transportprotokoll benötigten Angaben (Multicast-adressen etc.)

### Zeitinformation

Wie bereits angedeutet, kann eine Sitzung zeitlich beschränkt oder unbeschränkt sein. Beispiele dafür sind ein zeitlich begrenztes Telefongespräch oder ein ununterbrochen offener Chatroom. Es gibt jedoch auch noch eine Zwischenlösung. Sitzungen könnten jeweils nur zu bestimmten Zeiten aktiv sein. Deshalb ist es in SDP möglich eine Liste von Start- und Enddaten mitzugeben. Auch das periodische Vorhandensein einer Sitzung kann per SDP beschrieben werden. Zeitinformation wird jeweils zusammen mit der verwendeten Zeitzone angegeben. Diese Vorgehensweise ermöglicht Zeitangaben global eindeutig zu identifizieren.

### Öffentliche und private Sitzungen

Mit SDP ist es möglich, sowohl private als auch öffentliche Sitzungen zu beschreiben. Wenn eine Sitzung nur für bestimmte Personen offen sein soll, werden die SDP-Nachrichten Verschlüsselt. Da diese Informationen nur von berechtigten Nutzern eingesehen werden können, kann in verschlüsselten SDP-Nachrichten auch die verwendeten Verschlüsselungsinformationen der Mediendaten versendet werden.

### SDP Spezifikation

Die Sitzungsbeschreibung wird in Klartext geschrieben und besteht aus mehreren Zeilen. Jede Zeile beschreibt eine Eigenschaft. Die Zeilen haben die Form $<Typ>=<Wert>$ wobei $<Typ>$ jeweils aus einem einzelnen Buchstaben besteht. Bei der Typenangabe muss die Gross- und Kleinschreibung beachtet werden (T!=t). Der Inhalt von $<Wert>$ ist abhängig von $<Typ>$ und besteht aus einer strukturierten Zeichenfolge. Auch hier spielt die Gross-/Kleinschreibung eine Rolle. Neben dem '=' sind keine Leerzeichen erlaubt. In Tabelle 5.2 sind die verwendbaren Typenangaben beschrieben. Einzelne dieser Angaben müssen zwingend vorhanden sein, während andere wahlweise angegeben werden können. Die unverbindlichen Eigenschaften sind mit einem '*' versehen. In jeder SDP-Nachricht müssen jedoch alle Angaben in der genau gleichen Reihenfolge wie unten darstellt aufgeführt werden, egal ob ihnen Werte zugewiesen wurden oder nicht.

Ein Beispiel einer SDP-Beschreibung [4]:

```
v=0
o=mhandley 2890844526 2890842807 IN IP4 126.16.64.4
s=SDP Seminar
i=A Seminar on the session description protocol
u=http://www.cs.ucl.ac.uk/staff/M.Handley/sdp.03.ps
e=mjh@isi.edu (Mark Handley)
c=IN IP4 224.2.17.12/127
t=2873397496 2873404696
a=recvonly
m=audio 49170 RTP/AVP 0
```

```
m=video 51372 RTP/AVP 31
m=application 32416 udp wb
a=orient:portrait
```

Tabelle 5.2: SDP Parameter [4]

| **Sitzungsbeschreibung** | |
|---|---|
| **Typ** | **Wert** |
| v= | Protokollversion |
| o= | Besitzer und Sitzungsidentifizierer |
| s= | Sitzungsname |
| i=* | Sitzungsinformation |
| u=* | URI der Beschreibung |
| e=* | Emailadresse |
| p=* | Telefonnummer |
| c=* | Verbindungsinformation |
| b=* | Information über die Bandbreite |
| Eine oder mehrere Zeitinformationszeilen (siehe unten) | |
| z=* | Zeitzonen Anpassung (Bsp: +1,+2 etc.) |
| k=* | Schlüssel für die Chiffrierung |
| a=* | Keine oder mehr Sitzungsattributlinien |
| Keine oder mehr Medieninformationszeilen (siehe unten) | |
| **Zeitinformation** | |
| **Typ** | **Wert** |
| t= | Zeit wann die Sitzung aktiv ist |
| r=* | Keine oder mehr Wiederholungszeiten |
| **Medieninformation** | |
| **Typ** | **Wert** |
| m= | Medienname und Transportadresse |
| i=* | Medientitel |
| c=* | Verbindungsinformation |
| b=* | Information über die Bandbreite |
| k=* | Schlüssel für die Chiffrierung |
| a=* | Keine oder mehr Medienattributlinien |

### 5.2.4 Das RTP-Protokoll

Nachdem eine Verbindung per SIP aufgebaut und mit SDP beschrieben wurde, kann der Transfer der eigentlichen Information einer Sitzung beginnen. Diese Daten werden oft mit dem *Realtime Transport Protocol* übertragen. RTP kann entweder mit TCP- oder UDP-Paketen übertragen werden. Meistens werden UDP-Pakete verwendet, da bei TCP jedes Paket bestätigt werden muss. Bei Echtzeit-Multimedia-Anwendungen ist es oft weniger

wichtig, dass alle Pakete ankommen. Viel wichtiger ist, dass die meisten Pakete rechtzeitig ihr Ziel erreichen. RTP bietet Funktionen um Echtzeitdaten[1] in einem Ende-zu-Ende-Netzwerk zu übermitteln. Es werden keine Funktionen für die Reservation von Ressourcen implementiert und deshalb kann RTP keine Garantie für die Dienstgüte abgeben. Durch die Verwendung von *RTCP (RealTime Control Protocol)*, einem Protokoll welches die per RTP übermittelten Daten überwacht, werden einfachste Kontroll- und Identifikationsmechanismen implementiert. RTP und RTCP funktionieren unabhängig von den darunter liegenden Netzwerk- oder Transportprotokollen.

### 5.2.5 Ein einfaches Beispiel eines Verbindungsaufbaus

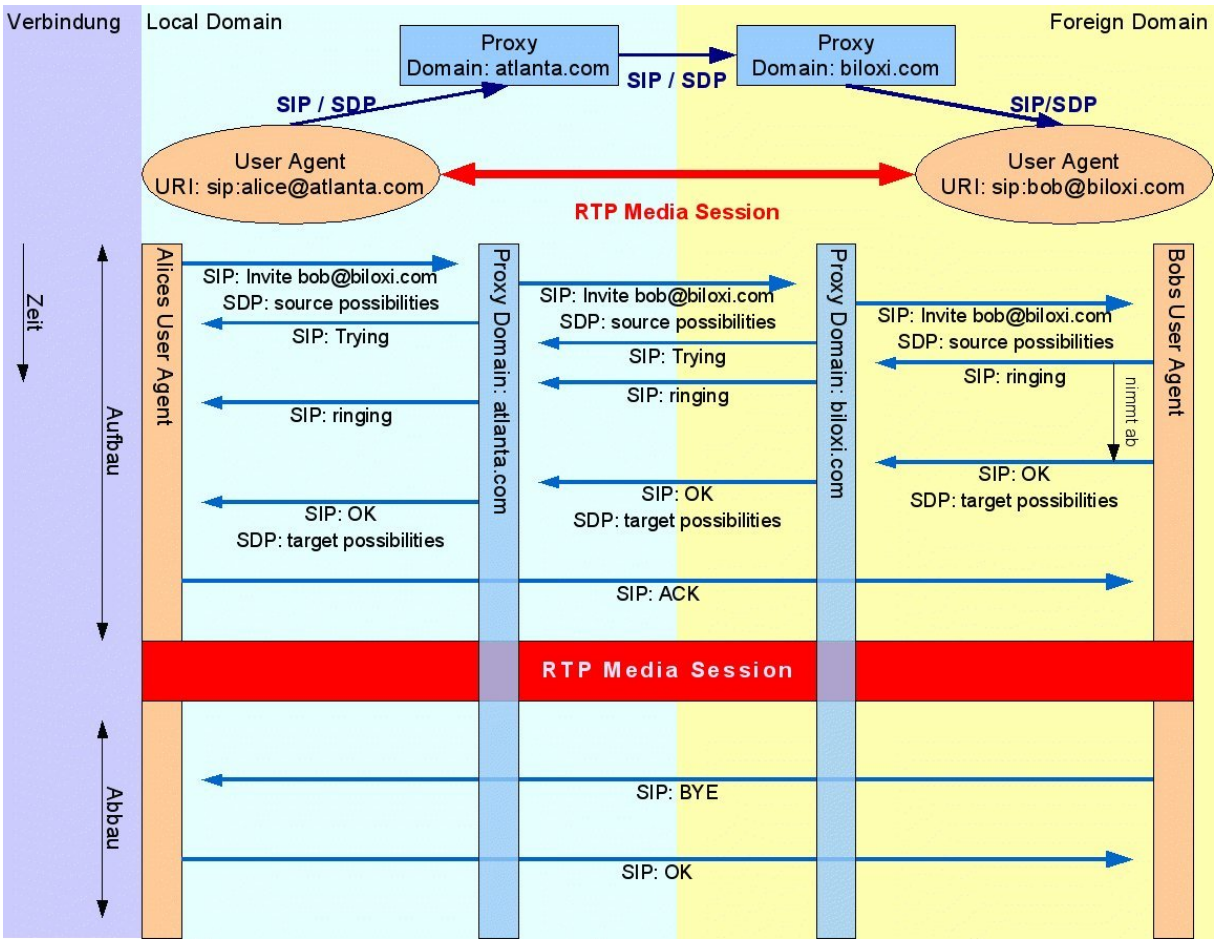


Abbildung 5.2: Einfacher Verbindungsaufbau

Um die Funktionsweise von SIP besser veranschaulichen zu können, werden wir im folgenden die grundlegenden Vorgänge eines SIP-Verbindungsaufbaus anhand eines Beispiels beschreiben. Die Teilnehmer der in Abbildung 5.2 gezeigten Verbindung sind Alice und Bob. Alice wird von ihrem Softphone[2] aus versuchen eine Verbindung mit dem SIP-Telefon

[1] Echtzeitdaten sind zum Beispiel Audio-, Video oder Simulationsdaten – Es ist wichtig dass der Empfänger diese in Echtzeit erhält und nicht zu spät.

[2] Ein Softphone ist ein durch eine Software simuliertes Telefon.

von Bob herzustellen. Jeder Teilnehmer ist bei einem SIP Anbieter registriert und ist dort unter einer SIP URI erreichbar. Alice ist in der Domäne atlanta.com zu Hause und erhält die Adresse sip:alice@atlanta.com während Bob unter der Adresse sip:bob@biloxi.com gefunden werden kann. Da die Teilnehmer natürlich nicht nur dann erreichbar sein sollen, wenn sie sich gerade in der Heimdomäne befinden, erfolgt der Verbindungsaufbau bei SIP über so genannte Proxyserver. Proxyserver nehmen Anfragen entgegen und geben diese für die anfragende Stelle weiter.

SIP funktioniert – ganz ähnlich wie HTTP – auf einem Anfrage-/Antwort-Modell. Jede Transaktion besteht aus einer Anfrage, die das Ausführen einer bestimmten Funktion auslöst sowie aus mindestens einer Antwort. In unserem Beispiel beginnt der Verbindungsaufbau mit einer INVITE-Anfrage. Da Alice nicht wissen kann, wo sich Bob zur Zeit gerade aufhält, schickt ihr Softphone diese Anfrage zuerst an den in ihrer Domäne zuständigen SIP Server. Welcher Server zuständig ist kann entweder in einer Konfigurationsdatei angegeben oder über DHCP-ähnliche Dienste angekündigt werden. In unserem Fall wäre dies atlanta.com. Alice erhält vom Proxyserver eine TRYING-Antwort und weiss nun, dass Ihre Anfrage angekommen ist und Bob gesucht wird. Der Proxyserver atlanta.com lokalisiert nun den Proxyserver von Bob z.B. über eine Art DNS Anfrage und erhält dadurch die IP Adresse von biloxi.com. Bevor die INVITE-Anfrage nun an den nächsten Proxyserver weitergeleitet wird, schreibt atlanta.com noch seine eigene Adresse in den Header der Nachricht. Dadurch wird sichergestellt, dass die Antworten auch wieder den gleichen Weg zum ursprünglichen Anrufer zurückfinden. Auch auf dieser Stufe reagiert biloxi.com mit einer TRYING-Antwort auf die INVITE-Nachricht von atlanta.com. Im nächsten Schritt muss nun die momentane IP Adresse von Bob gefunden werden. Dazu konsultiert biloxi.com eine Datenbank. Diese Datenbanken werden *location services* genannt. Auch der Proxyserver in biloxi.com fügt in die INVITE-Anfrage seine eigene Adresse ein und sendet darauf die Nachricht an das SIP Telefon von Bob. Darauf sendet das Telefon die RINGING-Antwort zurück und beginnt damit Bob auf den eingehenden Anruf aufmerksam zu machen (z.B. durch Läuten). Die vom Telefon gesendete Antwort wird nun über alle Proxystationen wieder zurück an Alice gesendet, wobei jeder Proxy seine Adresse wieder aus dem Header der Nachricht entfernt. Diese Technik verringert die Anzahl benötigter DNS-Anfragen erheblich, denn nur bei der initialen Suche nach Bobs Telefon müssen DNS Anfragen getätigt werden. Nachdem sich Bob dazu entschlossen hat, den Anruf entgegen zu nehmen, wird eine weitere Antwort (OK) an Alice geschickt, welche diese wiederum mit einer ACK-Nachricht quittiert. Danach steht die Verbindung und die Multimedia-Konversation kann beginnen. Beendet wird eine Sitzung, wenn einer der beiden Teilnehmer eine BYE-Nachricht schickt und diese vom Gegenüber mit einem OK beantwortet wird. Hierbei ist keine ACK-Nachricht notwendig.

### 5.2.6 Beispiel mit Umleitung

Eine Umleitung wird verwendet, um kurzzeitig seine Position wechseln zu können, ähnlich wie eine Telefon-Umleitung. Dazu wird auf dem Location-Server eine Umleitungs-URI hinterlegt, welche auf eine andere Domäne und Benutzer verweist. Dies bringt den Vorteil, dass man zum Beispiel einen Anruf auf einen Anrufbeantworter oder auf sein Mobil-Telefon umleiten kann. Und zwar funktioniert eine SIP-Umleitung, wie sie in der

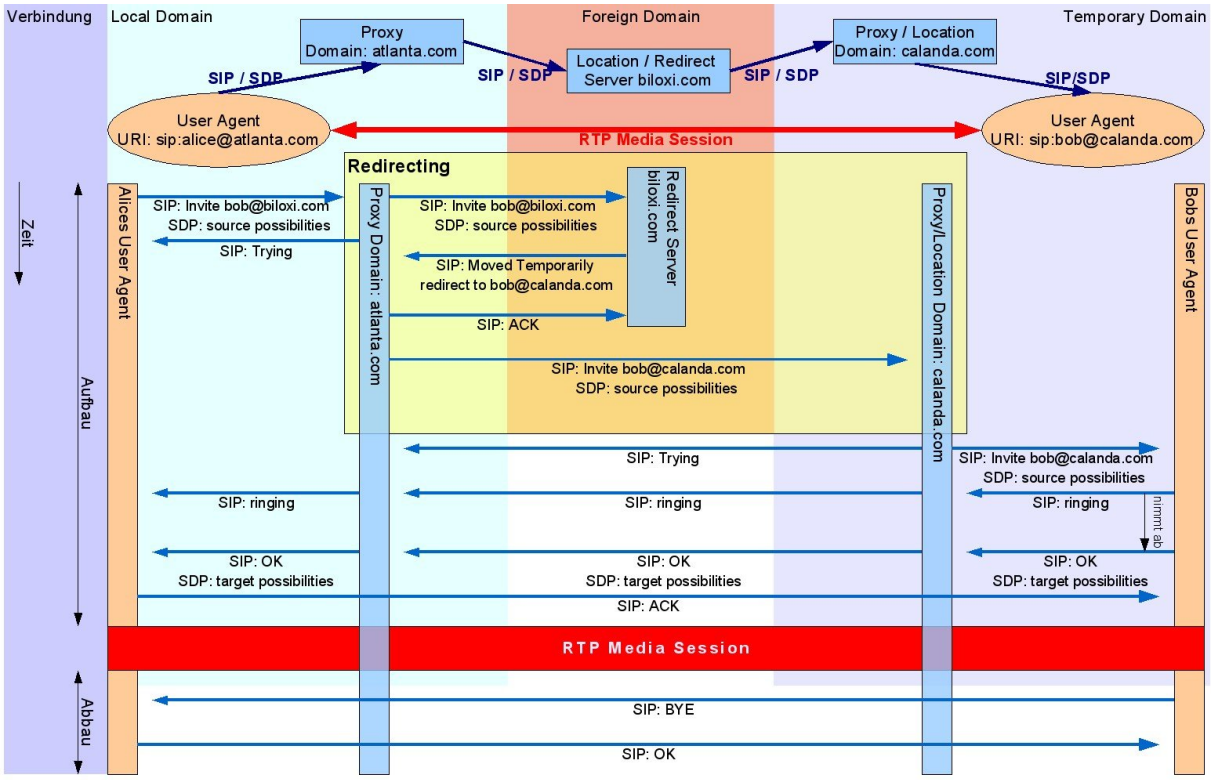


Abbildung 5.3: Verbindungsaufbau mit einem Redirect

Abbildung 5.3 dargestellt ist, folgendermassen: Wenn eine Verbindung aufgebaut werden soll, sendet in unserem Beispiel Alice eine Invite-Methode mit der gewünschten Ziel-URI an ihren Proxy. Dieser leitet die Anfrage weiter an den Proxy, der die gewünschte Ziel-Domäne (biloxi.com) verwaltet. Der Ziel-Domänen-Proxy startet eine Anfrage beim Location-Service, welcher die Umleitungs-URI gespeichert hat. Diese neue URI wird mittels REDIRECT-Anweisung an Alices Proxy-Server zurück gesendet.

Der Proxy-Server von Alice kann nun eine neue INVITE-Nachricht an den neuen Proxy-Server (calanda.com) senden, welcher für die neue URI (bob@calanda.com) zuständig ist. Bob ist nun erreichbar und sendet ein RINGING-Signal an Alice. Diese Antwort wird wiederum über alle beteiligten Proxies zurück gesendet. Sobald der angerufene Teilnehmer die Verbindungsanfrage akzeptiert, beispielsweise durch Abheben des Telefonhörers, wird analog zum RINGING-Signal eine OK-Nachricht über alle beteiligten Knoten zurückpropagiert. Der Ursprungsknoten erhält nun also die Bestätigung, dass der Zielknoten seine Einladung angenommen hat. Die anfragende Stelle quittiert dies wiederum mit dem Versand einer ACK-Nachricht, welche nun aber nicht mehr über die Proxies geleitet wird, da Alice nun Bobs neue IP-Adresse kennt. Die Datenverbindung kann nun, gleich wie im Beispiel ohne Umleitung, stattfinden.

## 5.3 Komponenten

Es gibt verschiedene Komponenten, welche benötigt werden um SIP zu verwenden. Die Kommunikation zwischen den einzelnen Komponenten läuft normalerweise über beliebige Ethernet-Verbindung.

### 5.3.1 Endgeräte

SIP wird verwendet, um Verbindungen zwischen mehreren Benutzern für multimediale Kommunikation zu ermöglichen. SIP wird vor allem im Voice-over-IP Segment eingesetzt. Die Endgeräte sind die letzte Stufe zum Benutzer. Im folgenden stellen wir die wichtigsten Hard- und Softwareendgeräte vor.

#### Hardphones

Abbildung 5.4: ProSIP SIP Telefon [24]

Hardphones sehen wie normale, alltägliche Telefone aus. Im Unterschied dazu werden die Daten nicht über das normale Telefon-Netz sondern per Internet mit der VoIP-Technologie versendet. Die Telefone bieten neben den gewohnten Funktionen auch eine Vielzahl von weiteren Möglichkeiten an. Der Vorteil solcher Hardphones ist, dass sie so wie ganz gewöhnliche Telefone zu bedienen sind. Die Benutzer müssen sich nicht an eine neue Technologie gewöhnen.

#### Softphones

Softphones (Bild 5.5) ermöglichen es einem Benutzer ein Telefon in einem Computer zu vereinen. Es wird kein Hardphone benötigt, wenn über das Internet telefoniert wird. Das Telefon kann als Programm im Computer simuliert werden und bietet die selben Möglichkeiten, wie ein Hardphone. Telefoniert wird sinnvollerweise über ein Headset. Es gibt mittlerweile auch Softphones für Pocket-PCs. Somit ist es möglich, überall unterwegs mit der benötigten Voice-over-IP Verbindung über Wireless LAN zu telefonieren.


Abbildung 5.5: EyeBeam SIP Softphone [23]

#### Chatclients

Chatclients funktionieren eigentlich wie Softphones, jedoch wird keine Sprache sondern Text übermittelt. In der Abbildung 5.5 ist ein Chatclient auch direkt in das Softphone integriert. Anstelle von Audio-Daten werden aber Text-Daten übermittelt.

#### Videoconferencing Software/Hardware

SIP-Geräte für Videokonferenzen haben die selben SIP-Funktionen, wie alle anderen SIP Endgeräte. Jedoch werden für die Video-Übertragung andere Bandbreiten benötigt. Softphones haben den Vorteil, dass sie alle verschiedenen Übertragungs-Möglichkeiten integriert haben können. So entfällt die Anschaffung von teurer Hardware. Je nach Anwendungswunsch, werden Kameras und/oder Mikrofone benötigt. Das obige Bild (Bild: 5.5) zeigt ein Telefon, welches auch ein Bild-Telefon integriert hat.

### 5.3.2 Proxies

Die SIP-Clients sind normalerweise auf einen SIP-Proxy konfiguriert, welcher die SIP-Verbindungen an die gewünschte Ziel-Adresse weiterleiten. Sie verhalten sich gleichzeitig als Client und als Server für Verbindungen. Ist einmal eine Verbindung aufgebaut, gehen alle SIP-Pakete über dieselbe Leitung hin und zurück. SIP-Proxies können auch verwendet werden, um Anfragen zu verändern. Beispielsweise kann ein Proxy mit Hilfe eines Registrars kontrollieren, ob ein Benutzer überhaupt berechtigt, ist einen Anruf zu senden. Einfachheitshalber ist der SIP-Proxy und der SIP-Registrar-Server meistens auf der selben Maschine installiert.

### 5.3.3 Registrare

Registrare verarbeiten die SIP-register-Methode. Die Benutzer melden sich bei den Registraren an, sie sind für die Benutzer-Verwaltung zuständig. Ein Registrar speichert für jeden Benutzer die jeweilige SIP-Adresse – zusammen mit der aktuellen IP-Adresse – auf

einem *Location-Service-Server*. Durch diese Information kann der registrierte Benutzer lokalisiert werden und einen SIP-Anruf empfangen.

### 5.3.4 Location Service

Mit *Location Service* wird die Datenbank bezeichnet, in welche die Registrare die Informations-Paare aus SIP-Adresse und IP-Adresse speichern. Diese Paare werden *Bindings* zu Deutsch *Verbindung* genannt. Die Datenbank welche diese *Bindings* beinhaltet beantwortet die von Proxyservern gestellten Anfragen. Als Abfrageprotokoll kommt zum Beispiel LDAP (Lightweight Directory Access Protocol) zur Anwendung.

### 5.3.5 Redirect Servers

Wenn ein der Ziel-Benutzer nicht gefunden wird, sendet der Redirect Server dem Quell-Benutzer die möglichen Orte für den Ziel-Benutzer. Der Quell-Client versucht dann einen neuen Verbindungs-Aufbau mit der neuen URI, welche er vom Redirect-Server bekommen hat.

## 5.4 Mobilität

Der Titel dieser Arbeit beinhaltet die beiden Wörter *mobile environment*. Nun stellt sich jedoch die Frage, was unter einem *mobilen Umfeld* zu verstehen ist. In der Informatik lässt sich der Begriff der Mobilität in vier Bereiche teilen:

1. Persönliche Mobilität
2. Session-Mobilität
3. Dienstemobilität
4. Endgerätemobilität

Im folgenden werden diese verschiedenen Typen von Mobilität genauer erläutert.

### 5.4.1 Persönliche Mobilität

Persönliche Mobilität bedeutet, dass es möglich ist, einen Benutzer generell unter einer global eindeutigen Adresse zu kontaktieren. In SIP ist dies implementiert, indem ein Benutzer sich von jedem *User Agent* aus bei seinem Registrar anmelden kann. Danach ist er an seinem aktuellen Aufenthaltsort unter seiner SIP-Adresse erreichbar. Bild 5.6 zeigt, wie Benutzer *User* sowohl von Ort *A* wie auch an Ort *B* nach dem jeweiligen REGISTER-Befehl unter *User@home* erreichbar ist.

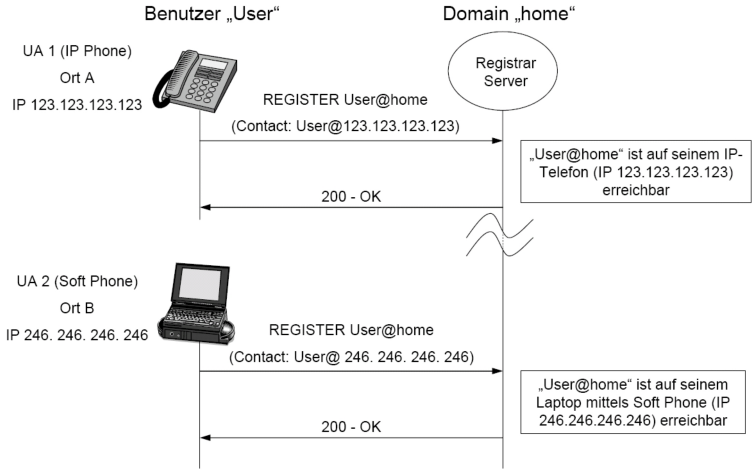


Abbildung 5.6: Persönliche Mobilität [22]

### 5.4.2 Session-Mobilität

Die Möglichkeit eine Verbindung oder Teile davon während einer Sitzung auf ein anderes Gerät zu übertragen, wird als *Session-Mobilität* bezeichnet. Zum Beispiel kann es während eines SIP-Gesprächs wünschenswert sein, das Gespräch von einem stationären Telefon auf ein Mobiltelefon zu übertragen oder nur die Videoübertragung von einem Gerät mit einem kleinen Bildschirm auf ein anderes mit einer grösseren Anzeige zu wechseln. SIP definiert derzeit zwei verschiedene Möglichkeiten, um dies zu realisieren. Beide funktionieren durch eine teilautomatisierte Vermittlung einer Verbindung über eine an der Kommunikation selbst unbeteiligte Stelle. Diese so genannte *Third-Party Call Control* kann entweder per INVITE-Befehl oder mittels einer REFER-Nachricht erreicht werden.

### 5.4.3 Dienstemobilität

Unter Dienstemobilität versteht man die Möglichkeit, für die Benutzer jederzeit und unabhängig vom aktuellen Aufenthaltsort auf Dienste seines SIP-Anbieters Zugriff nehmen zu können. Beispiele für solche Dienste sind Kontaktlisten, Telefonbeantworter, Konferenzschaltungen und dergleichen. Die Verfügbarkeit dieser Dienste hängt nicht vom verwendeten SIP-Endgerät ab. Auf diese Weise können auch Daten zwischen den benutzten User Agents synchronisiert werden. So ist denkbar, dass ein lokal zwischengespeichertes Adressbuch jeweils über den SIP-Server synchronisiert wird.

### 5.4.4 Endgerätemobilität

Wenn sich ein Benutzer während einer Multimediakommunikation frei bewegen und damit unter Umständen auch das aktuelle IP-Netz wechseln kann, ohne dass die Verbindung dabei unterbrochen wird, spricht man von Endgerätemobilität. Endgerätemobilität ist also eine Verschmelzung der *Persönlichen Mobilität* und der *Session-Mobilität*. Beim Wechsel

von einem Subnetz in ein anderes werden die in Talk 4 beschriebenen *Handovers* benötigt. Generell sind bei der Implementation der *Endgerätemobilität* drei verschiedene Situationen zu unterscheiden:

1. Verhalten im Zustand der Gesprächsbereitschaft: Um auch nach einem Subnetzwechsel für eine SIP-Kommunikation erreichbar zu bleiben, müssen allfällige IP-Adressen-Wechsel mit einer erneuten REGISTER-Nachricht an den zuständigen Registrar gemeldet werden. In Abbildung 5.7 wird dies veranschaulicht.

2. Verhalten während einer laufenden Sitzung: Falls während einer Verbindung das IP-Netz gewechselt wird, müssen die Kommunikationspartner über eine erneute INVITE-Nachricht die neue IP-Adresse erhalten. In der Zeitspanne zwischen Verlassen des alten Netzes und Reinitialisierung im neuen Netz kann es beim Nutzdatenaustausch zu einer kurzen Unterbrechung kommen. Dieses Verhalten wird in Abbildung 5.8 veranschaulicht.

3. Verhalten beim kurzfristigen Ausfall der Netzverbindung (z.B. durch ein Funkloch): Bei Unterbrechungen bis zu einer halben Minute werden bei SIP keine speziellen Funktionen gestartet. Die wartende Partei wiederholt einfach ihre Anfrage in der Annahme, dass die IP-Pakete auf dem Weg vom Sender zum Empfänger verloren gingen. Für Ausfälle über einen längeren Zeitraum bietet SIP die REDIRECT-Funktionalität. Wenn sich ein Benutzer in ein Funkloch begibt, wird der zuständige Proxy-/Redirectserver allen weiteren Anfragen die Antwort *302 - Moved temporarily* zurückschicken. Sobald der angerufene Benutzer wieder erreichbar ist, auch im Falle eines gleichzeitigen Netzwechsels, werden Anfragen wieder an den Benutzer weitergeleitet. INVITE-Nachrichten, welche nicht an ihr Ziel weitergeleitet werden können, verbleiben für eine gewisse Zeit im Speicher des SIP-Proxies. Nach Ablauf dieser Zeit werden noch hängige INVITE-Anfragen an einen nichterreichbaren Benutzer gelöscht. Danach muss erneut eine Einladung an den gewünschten Benutzer geschickt werden.

Wie die oben beschriebenen Situationen zeigen, implementiert SIP die wesentlichen Funktionen von Mobilität. SIP kann nicht nur in heterogenen Netzstrukturen sondern auch unabhängig von der verwendeten Netztechnologie (WLAN, UMTS, Ethernet, Tokenring etc.) eingesetzt werden. Zu beachten ist jedoch, dass nicht nur SIP sondern auch die RTP-Verbindungen über ähnlich flexible Möglichkeiten verfügen müssen. SIP kann wohl den Verbindungsaufbau und Unterbrüche über die bekannten Methoden realisieren, für eine brauchbare Kommunikationslösung sind jedoch auch auf Seiten des Nutzdatentransfers Flexibilität und Fehlertoleranz von Nöten.

## 5.5 Schwachstellen und Probleme von SIP

### 5.5.1 NAT und Firewalls

Ein wichtiger Bestandteil der heutigen Netzwerkinfrastruktur sind Firewalls. Sie schützen die Netzwerke von unberechtigten Zugriffen auf Netzwerkebene und ermöglichen per NAT

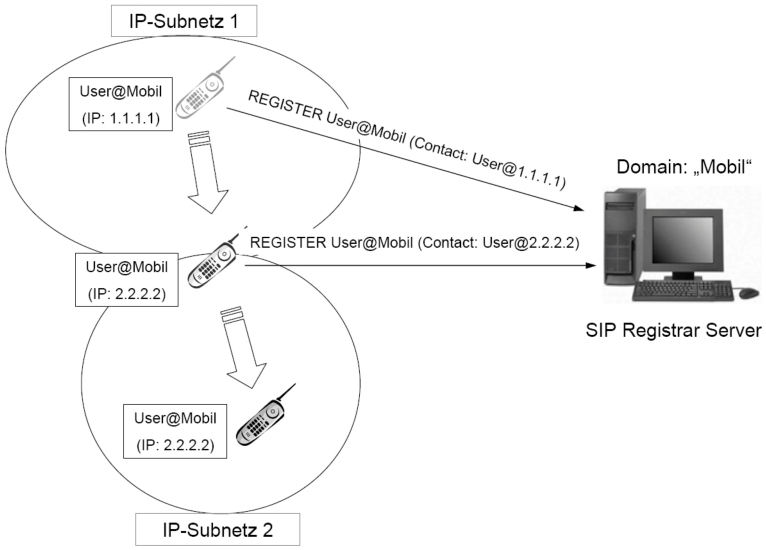


Abbildung 5.7: Erneutes Registrieren nach IP-Wechsel [22]

(Network Address Translation) mehreren Benutzern das gleichzeitige Verwenden einer einzigen Public-IP, einer im Internet eindeutigen öffentlichen IP-Adresse. Firewalls arbeiten typischerweise mit Filterregeln welche beschreiben, ob ein IP-Paket von Quelladresse A und Quellport X an eine Zieladresse B auf dem Zielport Y weitergeleitet werden soll.

Allgemein bekannte Anwendungsprotokolle wie Telnet oder SMTP arbeiten auf standardisierten Ports. Auch für den Verbindungsaufbau mit SIP existiert ein reservierter Port (5060). Somit kann eine Firewall so konfiguriert werden, dass SIP-Daten passieren können. Ein Problem ergibt sich jedoch mit den Multimediadaten, welche nach dem Verbindungsaufbau zwischen den Teilnehmern einer SIP-Kommunikation versendet werden müssen. Diese werden typischerweise auf dynamisch zugewiesenen Ports (>1023) versendet und deshalb kann einer Firewall keine eindeutige Regel für das Filtern dieser Daten einprogrammiert werden. Abhilfe schaffen hier sogenannte Stateful-Firewalls, welche sich den Verbindungszustand einer Kommunikation 'merken'. Dazu muss jedoch die Firewall den Inhalt von SIP-Nachrichten analysieren und zusätzlich Informationen über aufgebaute Verbindungen speichern können. Eine SIP-Sitzung kann also nur aufgebaut werden, wenn die im Netzwerk verwendeten Geräte auch SIP-kompatibel sind.

Es gibt jedoch noch ein weiteres Problem mit SIP und Firewalls: Die in den Headern von SIP-Nachrichten referenzierten IP-Adressen müssen nicht zwangsläufig öffentliche IP-Adressen sein. Wenn zum Beispiel Benutzer A in einem Intranet die Adresse 192.168.1.130 hat, kann er von aussen (ausserhalb des Intranets) nicht gefunden werden. Eine Lösung für diese Probleme bieten so genannte Applikation Level Gateways (ALG), welche die Header der verwendeten Protokolle auslesen. Im Fall von NAT werden dann die jeweiligen internen Adressen durch diejenige der Firewall ersetzt. Bei eingehenden Paketen werden die Zieladressen dann wieder in die jeweilige interne Zieladresse umgewandelt. Um diesen Dienst jedoch anbieten zu können, müssen die Firewalls das jeweilige Protokoll 'kennen'. ALGs, welche gängige Anwendungsprotokolle wie FTP, DNS oder HTTP beherrschen, gehören zur Standardausrüstung gängiger Firewalls. Es bleibt also zu hoffen, dass auch bald ALGs mit Unterstützung für SIP-Nachrichten ein Standardbestandteil von Netzwerkge-

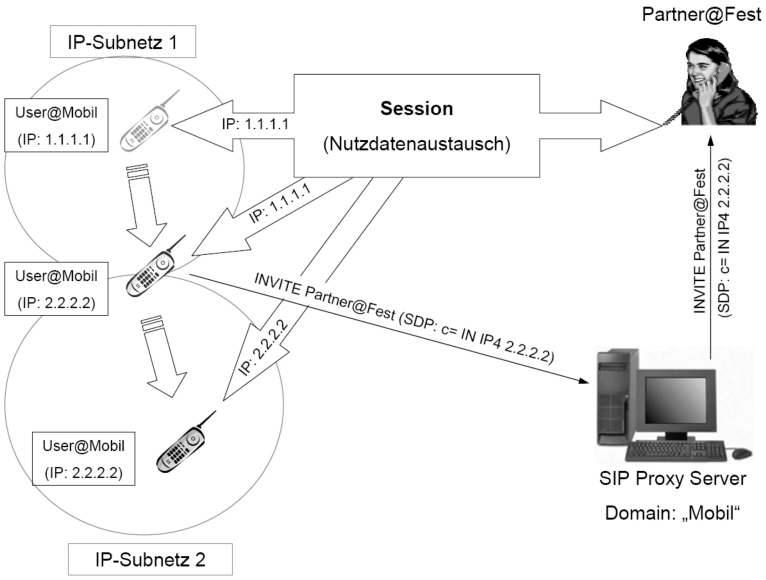


Abbildung 5.8: IP-Wechsel während Sitzung [22]

räten sein werden.

Eine weitere Möglichkeit, die Daten einer SIP-Kommunikation durch eine Firewall zu schleusen, bietet das Protokoll *Real Specific IP (RSIP) [25].* Bei der Verwendung von RSIP leiht sich der SIP-Client eine öffentliche IP des RSIP-Servers und baut eine Tunnel-Verbindung zu diesem auf. Alle ein- und ausgehenden SIP-Pakete werden dann durch diesen Tunnel verschickt. Somit ist der SIP-Client für eine gewisse Zeit über eine öffentliche IP im Internet erreichbar.

Neben den oben beschriebenen Methoden gibt es jedoch noch weitere Möglichkeiten um das so genannte *NAT Traversal Problem* in den Griff zu bekommen. Eine neuere Arbeit von J. Rosenberg dazu trägt den etwas langen Titel *Interactive Connectivity Establishment (ICE): A Methodology for Network Address Translator (NAT) Traversal for Multimedia Session Establishment Protocols* und steht als Internet-Draft[3] unter [14] zum Download bereit. ICE verwendet verschiedene Komponenten aus anderen Protokollen. Diverse Änderungen in bestehenden Protokollen, und damit in Programmen, sind der Preis für diese nach Rosenberg *... single solution which is flexible enough to work well in all situations* (aus dem Internet-Draft über ICE [14]).

### 5.5.2 QoS – Quality of Service

Unter *Quality of Service* versteht man ganz allgemein die Dienstgüte in Kommunikationsnetzen. Kriterien für diese Dienstgüte sind:

- Die zur Verfügung stehende Bandbreite. Gemeint ist damit die Anzahl Bits die pro Sekunde über eine Leitung übertragen werden. Diese Zahl beschränkt natürlich –

[3]Internet-Drafts sind Vorschläge und Arbeitsdokumente der Internet Engineering Task Force (IETF)

je nach Kompressionsstufe – die Anzahl Echtzeitverbindungen, die gleichzeitig mit akzeptabler Qualität über eine Leitung betrieben werden können.

- Die maximale Latenz[4]: Bei Echtzeitanwendungen gibt es eine maximale Zeit, während der auf ein Datenpaket gewartet werden kann. Es macht wenig Sinn ein zu spät angekommenes Paket im Nachhinein wieder in die bereits vergangene Kommunikation einzusetzen.

- Paketverlust (engl. Paketloss): Diese Messgrösse ergibt sich zum grössten Teil aus den obigen Kriterien. Wenn ein Paket zu lange auf dem Weg zwischen Quelle und Ziel ist, kann es sein, dass es wiederholt gesendet werden muss. Pakete gehen jedoch nicht nur durch zu lange Wegzeiten verloren, sondern auch durch Netze mit zu hoher Auslastung. Bei vielen Paketverlusten kann es zu hohen Latenzzeiten kommen.

Eine Möglichkeit um sicherzustellen, dass Echtzeitverbindungen auch bei hoher Netzlast benutzt werden können, ist die unterschiedliche Priorisierung vom Netzwerkverkehr. Sowohl in IPv6 [5] wie auch im immer noch verwendeten IPv4 [6] wurde das Markieren von Datenpaketen mit einer Prioritätsstufe spezifiziert. Jedoch unterstützen nur wenige der zur Zeit eingesetzen Routern eine Priorisierung des Datenverkehrs. Das uns heute bekannte Internet bietet eigentlich nur einen Dienst an und alle Pakete werden mit der gleichen Priorität behandelt. Diese Art von Netzwerkmodell wird als Best-Effort-Modell bezeichnet, da jedes Datenpaket so schnell wie möglich weitergeleitet wird. Bezüglich der Dienstgüte können diese Art von Netzen jedoch keine Garantien abgeben. Genau diese Garantien wären aber in manchen Fällen sehr wichtig: Es spielt zum Beispiel keine grosse Rolle, ob eine Email 2 Minuten früher oder später beim Empfänger eingeht, während es bei anderen Anwendungen von essenzieller Wichtigkeit ist, ob man nun eine oder 20 Sekunden Zeitverzögerung in Kauf nehmen muss. Unglücklicherweise ist vor allem das menschliche Gehör extrem empfindlich auf zeitliche Unterbrüche oder Verschiebungen. Bereits Verzögerungen von 200 Millisekunden fallen dem durchschnittlichen Menschen beim Telefonieren auf und werden als unangenehm empfunden. Bis also die Internettelefonie eine wirkliche alternative für das normale Telefon sein wird, muss für dieses Problem eine Lösung gefunden werden.

Da SIP sowohl über IPv6 als auch über IPv4 funktionieren soll, muss man sich bei der Implementierung von SIP Gedanken über die Dienstgüte des verwendeten Netzwerkes machen. In der Spezifikation von SIP selbst werden keine Angaben für die Priorisierung von Datenpaketen gemacht. Dies macht auch Sinn, denn der Verbindungsaufbau selbst ist nicht besonders empfindlich auf Verzögerungen. Eine hohe Dienstgüte ist vor allem bei einer auf den SIP-Verbindungsaufbau folgenden Multimediakommunikation wichtig.

Oft wird eine hohe Dienstgüte einfach durch das Vergrössern der Bandbreite eines Netzwerks erreicht. Dies ist jedoch nicht in allen Fällen möglich, sei es aus finanziellen oder technischen Gründen. Wenn man zum Beispiel in kabellosen Netzen über das Internet telefonieren möchte, kann es durchaus sein, dass gerade ein anderer Benutzer in der gleichen Zelle die verfügbare Bandbreite durch einen grossen Download oder andere ressourcenhungrige Anwendungen auslastet. Die kabellose Netzwerktechnologie hat in den letzten

[4]Die Latenzzeit ist der Zeitraum zwischen einer Aktion und dem Eintreten einer Reaktion

Jahren enorme Fortschritte gemacht, jedoch sind auch hier allzuhohe Durchsatzraten einfach nicht erreichbar.

Weitere Möglichkeiten die Dienstgüte durch eine Priorisierung des Datenverkehrs zu erreichen bieten verschiedene Techniken und Anwendungen, welche auf der Netzwerkebene arbeiten. Diese können jedoch nicht in jedem Fall angewendet werden: während es für die firmeninterne Kommunikation möglich wäre, das firmeneigene Netz für Echtzeitkommunikation zu optimieren, ist man für die Kommunikation mit der Aussenwelt meist auf andere Netze und Infrastrukturen angewiesen. Auf diese externen Datennetze (z.B. Internet) hat man meistens keinen oder nur beschränkt physischen Zugriff und kann somit auch keine Optimierung vornehmen.

Um über das Internet eine hohe *Quality of Service* zu erreichen, können verschiedene Dienste definiert und deren Pakete mit unterschiedlicher Priorität versendet werden. Die *IETF* hat dazu zwei Vorschläge gemacht:

1. Integrated Services (IntServ) [26]: Hierbei werden die zur Verfügung stehenden Ressourcen explizit verwaltet und eine bestimmte Leistung für Echtzeitanwendungen reserviert. Das Ziel die Ende-Zu-Ende-Verzögerungszeiten zu verkürzen wird mit dem Aufbauen von virtuellen Verbindungen realisiert. Die Ressourcen für diese Verbindungen müssen entlang des kompletten Pfades zwischen den Teilnehmern reserviert werden. Und hier liegt auch der grosse Nachteil dieser Lösung: Da jeder Router diese Funktionen beherrschen muss, ist ein grossflächiger Einsatz dieser Methode eher unwahrscheinlich. Um Ressourcen zu reservieren, verwendet IntServ das Protokoll RSVP (Ressource Reservation Protocol [27]).

2. Differentiated Services (DiffServ): Im Gegensatz zu *IntServ* verfolgt *DiffServ* einen verbindungslosen Ansatz, um damit eine bessere Skalierbarkeit realisieren zu können. Auch hier sollen Dienste mit verschiedenen Prioritäten belegt und demnach unterschiedlich behandelt werden. Jedoch werden bei DiffServ keine Ressourcen reserviert, sondern jedem Paket wird eine Prioritätsstufe zugeordnet. Die Router, welche die Pakete weiterleiten, entscheiden nun aufgrund so genannter PHB-Regeln bei allen eingehenden Paketen, welches Sie zuerst behandeln. Dieses Verfahren wird als *Per-Hop Behaviour (PHB)* bezeichnet.

Obwohl mit INTSERV sicherlich eine höhere Dienstgüte erreicht werden könnte, stellt doch deren Implementation ein grosses Problem dar. Im Gegensatz dazu garantiert DiffServ zwar nicht eine ganz so hohe QoS. Der grosse Vorteil liegt darin, dass nicht alle Netzwerkkomponenten mit den von DiffServ benötigten Regeln und Algorithmen ausgerüstet sein müssen.

### 5.5.3 Teilerfolge

Während die Trennung vom Verbindungsaufbau und der eigentlichen Multimediaverbindung mit den damit verbundenen Ressourcenreservationen Vorteile bei der Standardisierung bringt, können aber auch weniger erfreuliche Nebeneffekte entstehen: Die Tatsache,

dass ein SIP-Verbindungsaufbau zustandekommt, und weder Firewalls noch zu knappe Bandbreite dies verhindern, bedeutet noch nicht, dass die darauffolgende Echtzeitverbindung auch stattfinden kann. Einschränkungen durch zu knappe Netzwerkressourcen, Sicherheitseinstellungen oder dem Unvermögen die für die Verbindung nötigen Ressourcen zu reservieren, können eine durch SIP aufgebaute Verbindung wieder zum Erliegen bringen. Es ist möglich, dass ein Teilnehmer seinen Gesprächspartner zwar anrufen kann, dass jedoch, sobald die angerufene Person den Hörer abnimmt, kein Datentransfer stattfindet. Ein Lösungsvorschlag für dieses Problem sieht vor, der INVITE-Nachricht ein zusätzliches Feld im SDP-Paket mitzugeben, womit das SIP-Gerät der angerufenen Person den Befehl erhält, erst dann mit dem Klingeln zu beginnen, wenn genügend Ressourcen für die Verbindung reserviert wurden.

## 5.6 Sicherheit

Die Kommunikation über das Internet ist nur begrenzt sicher, ob das nun E-Mails oder Telefongespräche sind. Es stellt sich die Frage, wie viel Sicherheit wollen die einzelnen Benutzer haben. Die Möglichkeiten, mehr Sicherheit für SIP-Kommunikation zu erreichen, sind vergleichbar wie bei jeder anderen Kommunikation über das Internet. Ein Beispiel sind Firewalls oder Zugangs-Berechtigungen zu sensibler Hardware.

Im Vergleich zur herkömmlichen Telefonie sehen wir bei Voice-over-IP ein Problem, dass es keine zusätzliche Werkzeuge braucht um eine Leitung abzuhören, da Voice-over-IP übers Internet transportiert wird. Es gibt im Internet genügend Tools zum hacken verschiedenster Software, genauso könnten Laien Voice-over-IP hacken und Telefonate mithören. Deshalb sollten sicher Firewalls im Einsatz sein. Je sensiblere Daten versendet werden, desto mehr Wert sollte auf Sicherheit gelegt werden. Das heisst, dass Firmen oder sogar Regierungen andere Anforderungen an die Sicherheit stellen als ein privater Anwender.

Bei mobilen Wireless-Geräten ist schon alleine die Netzwerk-Verbindung ein Sicherheitsrisiko. Aber auch nur mit einer WEP-Verschlüsselung kann man es einem Angreifer mit nur wenig Aufwand relativ schwer machen. Wenn man aber ganz mobil sein möchte hat man unterwegs eventuell öffentliche Zugangs-Punkte, welche natürlich nicht verschlüsselt sind. Das muss aber jeder selbst wissen, ob er unverschlüsselt kommunizieren möchte. Grundsätzlich ist es gleich unsicher wie unverschlüsseltes wireless-surfen.

In einem Firmennetz mit SIP-Kommunikation kann mit Hilfe von Firewalls, VPN und Verschlüsselung für Wireless-Verbindungen die grösst-mögliche Sicherheit erreicht werden. Die Firewall sollte nur verschlüsselte Verbindungen vom SIP-Proxy zulassen und die Clients sollten selbst keine direkten Verbindungen mit der Aussenwelt aufbauen können. Die totale Sicherheit gibt es aber nicht.

Bei Multimedia Übertragungen mittels SIP kann unterschieden werden zwischen einem sicheren Kontroll-Kanal und einem sicheren Daten-Kanal. Wie schon vorher erklärt wurde ist SIP nur für die Verbindungs-Kontrolle verantwortlich. Wenn SIP sicher ist, heisst es noch nicht, dass die Daten auch sicher sind. Jedoch ist es sehr wahrscheinlich, dass die Kommunikations-Daten unsicher, wenn schon der Verbindungsaufbau unsicher ist.

### 5.6.1 Mögliche Angriffs-Arten

- Eavesdropping: Dies ist die Bezeichnung für ein einfaches Abhören einer Leitung. Solche Attacken sind kaum zu entdecken, da die Angreifer selbst keine Pakete versenden. Vor allem bei unverschlüsselten Wireless Netzwerken kann jeder unbemerkt mithören.

- Man in the Middle Attacken. Bei Man in the Middle Attacken leitet der Angreifer alle Pakete einer Verbindung über seinen Rechner. Dadurch hat er die Möglichkeit aktiv am Gespräch teilzunehmen, Daten abzuhören und abzufälschen.

- Denial Of Service Attacken: Bei einer so genannten DOS-Attacke, werden soviele falsche Verbindungen aufgebaut, dass der Server keine richtigen Verbindungen mehr erstellen kann. Dadurch wird das Netz unbrauchbar, es wird jedoch nichts abgehört, sondern nur die Server zum Absturz gebracht.

### 5.6.2 Secure SIP (SIPS)

Mittels Secure SIP [21] ist es möglich die Steuerdaten (die SIP Pakete) verschlüsselt zu übertragen. SIPS Verbindungen können mittels URI (Unified Ressource Identifier) identifiziert werden. Eine solche SIPS-URI lautet zum Beispiel: sips:sicher@voip.net.

SIPS funktioniert, indem bei allen Verbindungen zwischen zwei Geräten, das heisst auch zwischen allen Proxy-Servern eine verschlüsselte Leitung aufgebaut wird. Da SIPS also auf jedem SIP Knoten benötigt wird aber erst in wenigen Geräten unterstützt ist, ist es selten möglich eine SIPS Verbindung zu verwenden. Oft geht es nur dann, wenn beide Gesprächspartner in der gleichen Domäne angemeldet sind. SIPS ist aber immer mehr im kommen und wird auf praktisch allen neuen Komponenten unterstützt.

### 5.6.3 Secure SIPFon

Secure SIPFon ist eine Art Voice-over-IP Router, an den ein Voice-over-IP Endgerät angeschlossen ist. Es ist ein Java-Programm, welches für eine Diplomarbeit entwickelt wurde (siehe [21]). SIPFon ermöglicht eine verschlüsselte Daten-Verbindung zwischen zwei Endgeräten, wenn beide SIPFon installiert haben. Dies ist möglich, da der Datentransfer direkt zwischen den beiden Endgeräten erfolgt. Die Funktionsweise von SIPFon wird im folgenden Bild gezeigt (5.9).

SIPFon handelt beim Verbindungsaufbau dynamisch Schlüssel zwischen den zwei Partnern aus. Das Bild 5.10 zeigt, wie die Schlüssel ausgehandelt werden. Dieser Schlüssel wird für jede Verbindung neu ausgehandelt. Der darauf folgende Daten-Transfer wird dann mit diesem generierten Schlüssel nach dem AES (Advanced Encryption Standard) Standard verschlüsselt und gesendet.

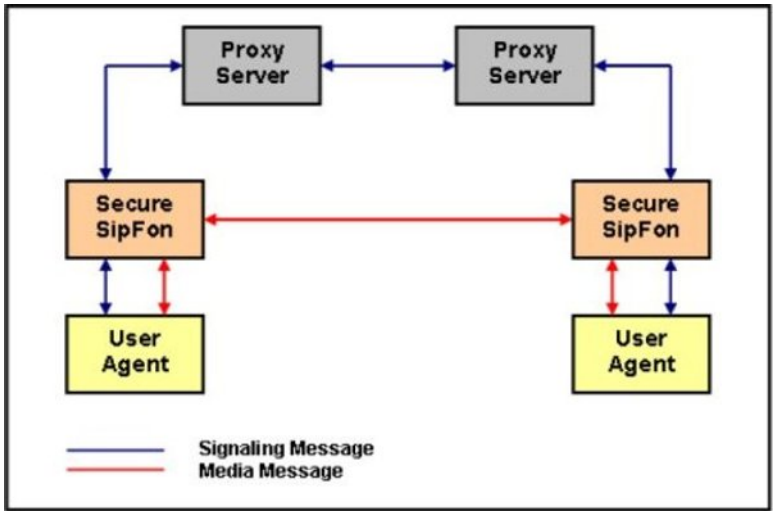


Abbildung 5.9: sicheres SIPFon Schema (Quelle: [21])

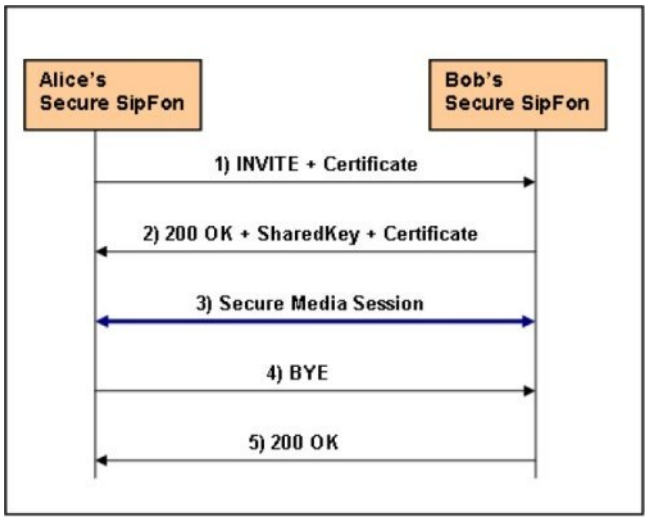


Abbildung 5.10: Der sichere SIPFon Schlüssel-Transfers (Quelle: [21])

### 5.6.4 Weitere Verschlüsselungsmöglichkeiten

Da Secure SIP nur bei den wenigsten Geräten funktioniert. Werden auch andere Verschlüsselungsmechanismen verwendet, so zum Beispiel IPSec VPN (Virtual Private Network) Verbindungen oder WPA (Wi-Fi Protected Access) in wireless Umgebungen. WPA ist für viele Anwender fast genau so sicher wie VPN und ist trotzdem um ein vielfaches günstiger. Vorallem bei Wireless-Kommunikation ist auf eine angemessene Verschlüsselung des Netzes nicht zu verzichten.

## 5.7 SIP in Aktion

SIP wird jeweils verwendet wenn ortsunabhängige und flexible Kommunikation benötigt ist. Es gibt jede Menge Anwendungs- Möglichkeiten. SIP ist ein offener Standard und

von jedem zu gebrauchen. Es gibt auch verschiedenste Open-Source Projekte, die SIP für Endbenutzer verwendbar machen.

Da SIP bloss eine Verbindung zwischen mehreren Teilnehmern aufbaut, ist es möglich über diesen Kanal beliebige Daten zu transferieren. Hauptsächlich wird im Moment SIP in der Voice-over-IP Kommunikation verwendet. Da jedoch Voice-over-IP besser sein soll als die herkömmliche Telefonie, sind SIP-Telefone oft gleichzeitig auch für Video-Konferenzen zu gebrauchen und bieten einen Chat-Client.

Weitere Anwendungen sind in Online-Spielen zu finden.

### 5.7.1 Universität Saarbrücken

An der Uni Saarland [11] wurde im Jahr 2000 ein Feldtest durchgeführt um die Verwendbarkeit von Voice-over-IP zu testen. Ab dem Jahr 2003 wurde für die interne Telefonie Voice-over-IP verwendet und das bestehende Telefonie-Netz sollte bald abgelöst werden können.

Folgende Anforderungen wurden an die Kommunikation gestellt:

- audiovisuell
- kostengünstig
- Ortsunabhängigkeit

Dank SIP konnte die Ortsunabhängigkeit erreicht werden. Die Hardware musste nur für die Server neu angeschafft werden um audiovisuell kommunizieren zu können. Die gesamte Kommunikation konnte aber auf den bestehenden Daten-Leitungen erfolgen und die Clients konnten per Softphone realisiert werden. Da SIP ein offener Standard ist, konnte das Projekt mittels günstiger Opensource-Software realisiert werden.

Die Studierenden der Uni Saarbrücken können nun mittels VPN-Zugang und SIP gratis telefonieren, vorerst jedoch nur mit umliegenden Universitäten.

### 5.7.2 SIP bei Siemens

Siemens hat rund um SIP ein eigenes Portal auf ihrer Webseite erstellt (siehe auch [18]) und bietet ihren Kunden eine weite Palette an Produkten welche auf dem SIP Protokoll basieren. Für die Bemühungen von Siemens ihren Kunden ein umfangreiches Angebot an SIP-basierte Produkten anbieten zu können wurde die Firma im Jahr 2005 zum zweiten mal in Folge mit einem Preis der *Frost&Sullivan-Awards* im Bereich *Technology Leadership* ausgezeichnet.

### 5.7.3 Agenten, Adressbücher, Mobil- und Festnetztelefonie

Da SIP ein offener Standard ist, ist es jedem möglich, eigene Anwendungen mit SIP-Unterstützung zu entwickeln. Dies wird zusätzlich dadurch gefördert, dass SIP auf bereits bestehenden und ebenso offenen Standards aufbaut. Durch die Verwendung des Internet können Dienste bereitgestellt, werden welche jene Informationen die bereits in einer digitalisierten Form vorliegen, mit den Diensten einer Telefonanlage verbinden. So ist es zum Beispiel möglich, die Telefonnummer (SIP Adresse eines Arbeitskollegen) in der Firmeninternen Datenbank zu suchen und den Verbindungsaufbau durch Klicken auf einen Hyperlink zu starten. Umleitungen auf Telefonbeantworter sowie Konferenzschaltungen sind natürlich wie gewohnt möglich. Information darüber, ob sich der gewünschte Gesprächspartner gerade an einem Ort befindet, wo er einen allfälligen Anruf entgegennehmen kann, könnte gleich in den Kontaktinformationen der Adresskartei dargestellt werden. Diese Funktionalität hat sich mit den sich immer mehr verbreitenden Instantmessenger-Systemen[5] etabliert.

Konfigurierbare Agenten stehen wie ein herkömmlicher Telefonbeantworter rund um die Uhr für die Benutzer im Einsatz. Zusätzlich zu der bekannten Bandfunktion, können diese Agenten jedoch noch einiges mehr: Der SIP-Benutzer gibt an, wann und wo er erreichbar sein möchte und eingehende Anrufe werden jeweils an den aktuellen Standort weitergeleitet. Dabei spielt es keine Rolle, ob sich die zu findende Person gerade in der Nähe eines Computers, eines SIP- oder sonstigen Mobiltelefons befindet. Anrufe können auch ins herkömmliche Festnetz durchgestellt werden.

Da die SIP-Geräte stets auch mit dem Internet verbunden sind, können kurze Mitteilungen – wie zum Beispiel Börsenkurse oder Nachrichten von Geschäftspartnern – direkt auf dem Display dargestellt werden.

## 5.8 Fazit und Ausblick

### 5.8.1 Vor- und Nachteile

Vorteile:

- Mobilität: sobald man eine Verbindung zum SIP-Server aufbauen kann, ist man erreichbar. Egal wo auf der Welt man sich befindet.
- Flexibilität: SIP-Server und Netze sind relativ einfach zu konfigurieren und an individuelle Wünsche anzupassen. Das gesamte Telefon-Netz läuft über meist bestehende Ethernet-Verbindungen. Daher braucht es beispielsweise für einen neuen Telefonanschluss keine Extra-Kabel.

[5] Instant Messenger sind kleine Programme mit denen man jederzeit mit Arbeitskollegen oder Freunden einen Chat beginnen kann.

- Kosten: Da die Internet-Leitungen immer schneller werden, kann man die bestehenden Bandbreiten für die Telefonie verwenden. Die Software beruht auf offenen Standarten und es gibt jede Menge Open-Source-Lösungen. Dadurch wird es möglich sein, proprietäre Telekommunikations-Firmen für die Telefonie zu umgehen und seine eigene Telefonanlage aufzubauen.

Nachteile:

- Qualität: Um angenehm zu telefonieren oder gar Videos zu übermitteln, muss eine gewisse Bandbreite garantiert werden können.
- Notfälle: Bei einem Stromausfall funktioniert das herkömmliche Telefonnetz meist noch, bei Voice-over-IP stehen die Chancen aber schlecht. Im weiteren sind Notrufe nur bei wenigen Anbietern unterstützt.

### 5.8.2 Ausblick

Seit den ersten Überlegungen und Arbeiten zu SIP ist schon reichlich Zeit vergangen. Dies kam dem Reifeprozess dieser Idee zu Gute. Diverse grössere und kleinere Unternehmungen haben damit begonnen, neue Produkte mit SIP-Kompatibilität in ihr Angebot aufzunehmen. Durch die wachsende Verbreitung von SIP steigen immer mehr Soft- und Hardwareproduzenten (z.B. Cisco, Nokia, Microsoft etc.) – zumindest teilweise – auf SIP um. Da SIP ein offener Standard ist, besteht die Hoffnung, dass die weiteren Entwicklungen rund um SIP nicht von einer einzigen Interessensgruppe diktiert wird, sondern, dass auch weiterhin öffentlich und kreativ an SIP weiterentwickelt wird. Dies ist in einer global vernetzten Umgebung von enormem Wert, da ein Standard der weltweit Verwendung finden könnte, auch viele verschiedene Bedürfnisse abdecken muss. SIP ist, obwohl sehr einfach und intuitiv aufgebaut, sehr flexibel anwend- und erweiterbar.

Natürlich ist die Weiterentwicklung von neuen Technologien wie Voice-over-IP oder Videokonferenzen nicht nur von der Weiterentwicklung von SIP abhängig. Gerade für diese Art von Anwendungen ist SIP jedoch bis heute eine zentrale Komponente und hat bereits bestehende Protokolle bezüglich Anwendbarkeit und Einfachheit weit hinter sich gelassen.

# Literaturverzeichnis


[1] Henning Schulzrinne: Session Initiation Protocol (SIP) Homepage, http://www.cs.columbia.edu/sip/, Besucht im April 2005

[2] Rosenberg et. al.: 'SIP: Session Initiation Protocol' ,RFC 3261, June 2002.

[3] Handley et. al.: 'SIP: Session Initiation Protocol' (RFC 2543), The Internet Society, http://www.faqs.org/ftp/rfc/rfc2543.txt, March 1999

[4] M. Handley, V. Jacobson: 'SDP: Session Description Protocol' ,RFC 2327, April 1998.

[5] S. Deering, R. Hinden: Internet Protocol, Version 6 (IPv6) Specification' ,RFC 2460, Dezember 1998.

[6] Postel, J.: 'Internet Protocol', RFC 791, September 1981.

[7] The Session Initiation Protocol (SIP) working group: 'Session Initiation Protocol' (Charta), http://www.ietf.org/html.charters/sip-charter.html, Besucht im April 2005

[8] H. Schulzrinne et. al.: 'RTP: A Transport Protocol for Real-Time Applications', RFC3550, July 2003.

[9] Ubiquity Software: Understanding SIP – Today's Hottest Communications Protocol Comes of Age, Ubiquity Software (2004)

[10] Fredrik Thernelius: SIP, Nat, and Firewalls - Master's thesis, Department of Teleinformatics, Kungl Tekniska Högskolan, May 2000.

[11] Homepage des Rechenzentrums der Universität Saarland, http://www.rz.uni-saarland.de/projekte/VoIP/, Edgar Scherer (2004), Besucht im April 2005

[12] Wiki über SIP, http://voip-info.org/wiki-SIP, verschiedene Autoren, Besucht im April 2005

[13] Wiki über SDP, http://voip-info.org/wiki-SDP, verschiedene Autoren, Besucht im April 2005

[14] J. Rosenberg: Interactive Connectivity Establishment (ICE): A Methodology for Network Address Translator (NAT) Traversal for Multimedia Session Establishment Protocols, The Internet Society, http://www.ietf.org/internet-drafts/draft-ietf-mmusic-ice-04.txt, February 2005

[15] Deutsche Wikipedia: Die freie Enzyklopädie, http://de.wikipedia.org, Besucht im Mai 2005

[16] Jonathan D. Rosenberg & Richard Shockey: 'The Session Initiation Protocol (SIP): A Key Component for Internet Telephony', http://www.cconvergence.com/shared/article/showArticle.jhtml?articleId=8700868, June 2000, Besucht im Mai 2005

[17] Thomas Kappler: Proseminar Rund ums Internet – Quality of Service und Internet-Telefonie, http://www.stud.uni-karlsruhe.de/ uphr/documents/qos-proseminar/ausarbeitung.html, Besucht im April 2005

[18] Siemens AG: http://networks.siemens.de/telefonanlagen/solutionprovider/home/siemens-und-sip.html, Besucht im Mai 2005

[19] Siemens AG: http://networks.siemens.de/telefonanlagen/solutionprovider/home/vorteile-von-sip.html, Besucht im Mai 2005

[20] Johannes Jakob: VoIP: SIP The Session Initiation Protocol, http://wwwbode.cs.tum.edu/zope/lectures/seminars/WS04/PSTK/doc/pstk-ws04-041220-sip-jakob-ausarbeitung.pdf, Dezember 2004, besucht im Mai 2005

[21] Hannah Lee: Securing IP Telephony: Secure SipFon, http://www.tu-harburg.de/ sx-hl0490/ssf/report.pdf, November 2004, besucht im Mai 2005

[22] Frank Weber: Mobilität und SIP, http://www.e-technik.org/aufsaetze_vortraege/ aufsaetze/weber_trick_itg_mobilfunk_6_04.pdf, November 2004, besucht im Mai 2005

[23] Xten Homepage, http://www.xten.com, besucht im Mai 2005

[24] ProSIP Homepage, http://www.prosip.com, besucht im Mai 2005

[25] Fredrik Thernelius: SIP, NAT, and Firewalls, http://www.cs.columbia.edu/sip/ drafts/Ther0005_SIP.pdf, Besucht im April 2005

[26] IETF: Integrated Services (intserv), http://www.ietf.org/html.charters/intserv-charter.html, Besucht im Mai 2005

[27] R. Braden: 'Resource ReSerVation Protocol', RFC 2205, September 1997.

# Kapitel 6

# Verfahren zur Umsetzung von mobilen DRM-Systemen

*Manuel Feier, David Holzer*

*Diese Arbeit versucht einen Überblick über die bisherige Entwicklung, aktuelle Standards und Technologien sowie einen Ausblick auf die kommenden mobilen DRM-Systeme zu schaffen. Anhand wirtschaftlicher Überlegungen und rechtlicher Restriktionen werden die Anforderungen an ein (Mobile) DRM System skizziert und mit OMA DRM ein konkretes Beispiel einer zukunftsträchtigen Mobile DRM Architektur im Detail betrachtet.Von der rechtlichen Seite werden vor allem Fragen des Urheberschutzes sowie der privaten Nutzungsrechte beleuchtet. Die wirtschaftlichen Aspekte konzentieren sich auf die Begründung der Notwendigkeit von Mobile DRM sowie die Entwicklung neuer Geschäftsmodelle und Distributionswege. Es wird ein kurzer Überblick über die aktuelle Marktsitation geboten sowie einige ausgewählte DRM-Systeme anhand des Anforderungskataloges eingeordnet. Abschliessend werden verschiedene Kritikpunkte zu diesem zweifellos kontroversen Thema dargelegt.*

## Inhaltsverzeichnis

## 6.1 Einführung in DRM, Grundbegriffe

In diesem Abschnitt werden die Grundbegriffe von 'Digital Rights Management' (DRM) erläutert und die Notwendigkeit für Mobile DRM aus Sicht der Industrie dargelegt. Es wird zudem aufgezeigt, welches die aktuellen und zukünftigen Einsatzgebiete für Digital Rights Management sind. Der Comic (Abbildung 6.1) nimmt Vorweg, was in folgenden Abschnitten erläutert wird.

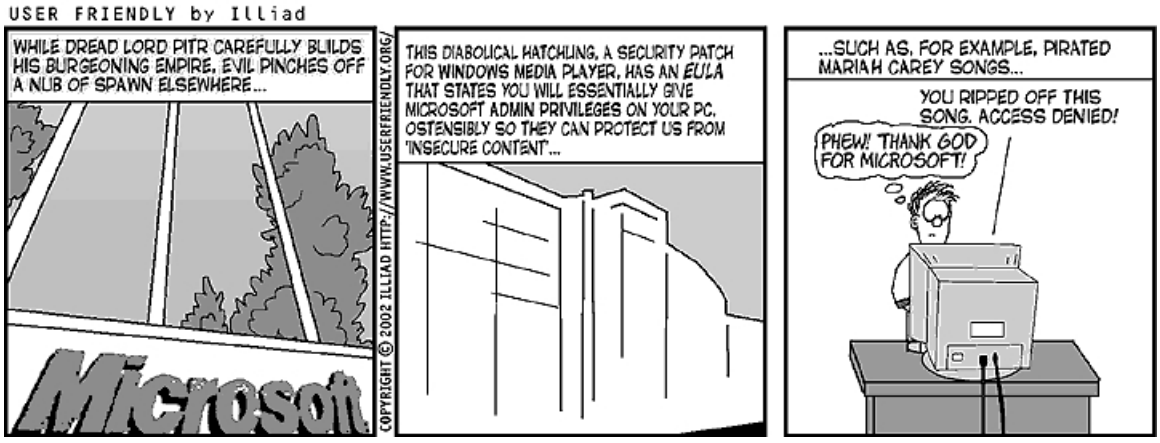


Abbildung 6.1: Comic 'Thank god for microsoft'

### 6.1.1 Begriffsdefinition

Der Begriff 'DRM' ist eine Abkürzung für 'Digital Rights Management', was wiederum ein Sammelbegriff für unterschiedlichste technische Verfahren ist, welche dem Schutz von immateriellen Inhalten durch seine Urheber bzw. Herausgeber dienen. Der Begriff 'Mobile DRM' (zu deutsch etwa: 'Mobiles Rechtemanagement') wird in der Fachliteratur hauptsächlich für folgende Eigenschaften eines DRM-Systemes verwendet:

Einerseits versteht man darunter jene Technologien und Verfahren, welche die oben erwähnten Schutzfunktionen vor allem in Hinblick auf mobile Endgeräte wie Handies, portable Musikabspielgeräte usw. Gewährleisten sollen; andererseits umschreibt der Begriff jene Verfahren, die es ermöglichen, geschützte Inhalte zwischen ensprechenden Endgeräten unter Beibehaltung der Rechte-Eigenschaften zu transferien. Die Anforderungen an mobile DRM sind identisch oder umfangreicher als diejenigen von herkömmlichen DRM Systemen (welche sich beispielsweise noch nicht einmal auf dem PC durchgesetzt haben). Zusätzlich zu den Problemen bei stationärem DRM kommt hier die ganze zusätzliche Problematik zum tragen, welche mobile Geräte mit sich bringen; zu nennen sind hier beispielsweise begrenzte Speicher-, Strom- und Rechenkapazität, erhöhte Komplexität bei der Datenübertragung sowie Einbussen bei der Verbindungsqualität.

### 6.1.2 Notwendigkeit für mobiles digitales Rechtemanagement

Die Notwendigkeit für (mobiles) Digital Rights Management ist keine technische, sondern eine wirtschaftliche. Kernpunkt der Problematik ist die Tatsache, dass digitale Daten heutzutage schnell, beliebig, praktisch kostenlos und ohne Qualitätsverlust über grosse Distanzen hinweg vervielfältigt werden können. Die früher aufgrund der physikalischen Eigenschaften der analogen Medien wie Schallplatten, Videokasetten usw. automatisch vorhandenen Schutzmechanismen gegen Vervielfältigung (Qualitätsverlust) sind bei den heutigen, digitalen Medien nicht mehr vorhanden. Obwohl sich der Massenmarkt auch mit digitalen, ungeschützten Medien (Audio-CD) weiterentwickelte, sieht sich die Medienindustrie (insbesondere die internationale Film- und Musikindustrie) durch die zunehmende Vernetzung der Gesellschaft in ihrer wirtschaftlichen Zukunft bedroht. Dieser Situation soll mit Digital Rights Management entgegengetreten werden.

Bei mobile DRM ist die technische Verhinderung der (möglicherweise illegalen) Kopiermöglichkeit aber nur ein Teilaspekt. Von ebenfalls in DRM-Systemen vorgesehenen Nutzungsregelungen auf digitale Inhalte erhofft sich die Industrie die Entwicklung von neuen Geschäftsfeldern und Distributionswegen. Die Inhalte – beispielsweise der Mobilfunkbranche – waren bisher relativ unaufwändige und 'billige' Produkte für das Mobiltelefon als Endgerät wie Logos, Animationen und alle Arten von Klingeltönen. Aufgrund des geringen Wertes waren für die Distribution solcher Inhalte bisher keine oder nur sehr einfache Kopierschutzmassnahmen ('Forward lock') erforderlich, nicht zuletzt weil die Inhalte oft für spezifische Gerätemodelle erstellt wurden und somit die illegale Verbreitung in der heterogenen Gerätelandschaft automatisch erschwert wurde.

Mit der stetigen Erhöhung der Bandbreite im stationären Internetbereich und der entsprechenden Entwicklung erster Geschäftsmodelle für digitale Inhalte (z.B. Apple iTunes Music Store) einerseits und der erwarteten Kapazitätssteigerung durch 3G-Technologien wie UMTS im Mobilfunkbereich andererseits steigt besonders im mobilen Markt das Bedürfnis nach branchenübergreifenden Lösungen zur Distribution von digitalen Inhalten.

Mit dem Ziel, Musik und Filme auf's Handy zu bringen, wird ein solides Rechtemanagement erforderlich, um die (eher konservativen) Inhaltslieferanten zur Distribution auf den neuen Kanälen zu bewegen. Apple ist dies mit iTunes im Internetbereich gelungen. DRM wird also nicht nur als Kopierschutzmassnahme betrachtet, sondern auch als Chance für die Etablierung von neuen Geschäfts- und Distributionsmodellen.

Eines dieser neuen Geschäftsmodelle ist beispielsweise die sogenannte 'Superdistribution' (zuweilen auch 'viral marketing' genannt). Die Grundidee dabei ist, dass die Verteilung von digitalen Inhalten von deren Nutzungsmöglichkeiten entkoppelt wird, wie Brad Cox für's Wired Magazin schrieb:[15]

> [..]superdistribution actively encourages free distribution of information-age goods via any distribution mechanism imaginable. It invites users to download superdistribution software from networks, to give it away to their friends, or to send it as junk mail to people they've never met.

Die Erklärung, weshalb dies wirtschaftlich sinnvoll sein sollte, wird so formuliert:

> Why this generosity? Because the software is actually meterware. It has strings attached, whose effect is to decouple revenue collection from the way the software was distributed. Superdistribution software contains embedded instructions that make it useless except on machines that are equipped for this new kind of revenue collection.

Die verteilten Inhalte sind also (dank DRM) schlussendlich nur auf Systemen, welche auch dafür bezahlen, voll nutzbar. Jedoch kann durch die kostenlose und grossflächige Verteilung die Hemmschwelle zum Kauf gesenkt werden, da die Inhalte ja bereits auf dem Endgerät liegen und quasi nur noch freigeschaltet werden müssen.

Die Abbildung 6.2 verdeutlicht die vielfältigen Aspekte, welche von DRM-Systemen abgedeckt werden müssen.

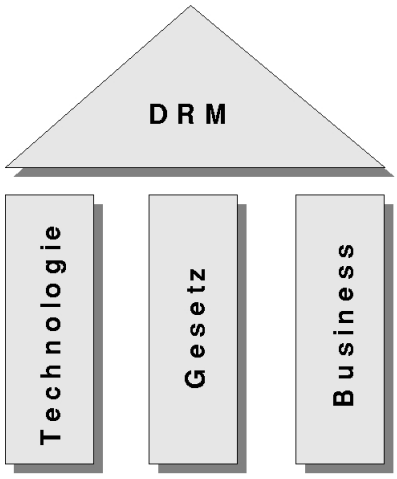


Abbildung 6.2: Die verschiedenen Aspekte von DRM

### 6.1.3 Bisherige und neue Einsatzgebiete

Bereits seit langem existieren Verfahren zur technischen Einschränkung der Verwendungsmöglichkeiten digitaler Inhalte. Zu erwähnen sind hier etwa die teilweise seit Jahrzehnten unternommenen Anstrengungen zum Kopierschutz, unter anderem auf folgenden Gebieten:

- Kopierschutz von Computerspielen (Diverse Kopierschutzverfahren)
- Kopierschutz von Anwendungssoftware (Diverse Kopierschutzverfahren)
- Kopierschutz von Filmen (DVD mit CSS, Macrovision usw.)
- Kopierschutz von Musik-CDs (Verschiedene Kopierschutzverfahren)

Diese Kopierschutzmechanismen entsprechen dem heutigen Verständnis von Mobile DRM jedoch nur teilweise, da sie allesamt nur der Verhinderung von Vervielfältigung dienen,

nicht aber der Durchsetzung von weitergehenden Nutzungsbeschränkungen, wie es heutige DRM-Systeme vorsehen. Wirklich an Bedeutung gewonnen hat (mobile) DRM erst in den letzten Jahren, als die schnelle Entwicklung des Internets die Tragweite, Geschwindigkeit und Einfachheit des Kopierens digitaler Inhalte schlagartig multiplizierte und damit zu einer realen Bedrohung für das Geschäft der Produzenten entsprechender Inhalte wurde. In dieser Ausgangslage hat es beispielsweise die Firma Apple mit ihrem Produkt 'iTunes' geschafft, einen marktfähigen Musikshop im Internet zu etablieren, wo Kunden Musik auf legale Weise erwerben und downloaden können. Die Verwendung des Apple-proprietären DRM-Systemes 'FairPlay' dürfte massgeblich dazu beigetragen haben, dass die Musiklabels Ihre Inhalte offiziell auf iTunes zum Download zur Verfügung stellen. Aber nicht nur im Bereich 'mobile music' sind interessante Entwicklungen zu beobachten. So baut die Firma Adobe, Hersteller der Acrobat-PDF-Dokumentprodukte, derzeit die DRM-Fähigkeiten ihres PDF-Formates aus, welche dem Autor zunehmende Möglichkeiten im Bereich der Rechteverwaltung seiner Dokumente ermöglicht.

## 6.2 Rechtlicher Rahmen

In diesem Teil soll der Rechtliche Rahmen beleuchtet werden, in dem DRM steht. Besonderes Augenmerk wird dabei auf die Situation in der Schweiz gelegt. Weiter wird die Lage in der restlichen Welt beleuchtet, wobei vor allem die Situation in den Vereinigten Staaten von Amerika und der Europäischen Union betrachtet wird.

### 6.2.1 Rechtliche Situation in der Schweiz

Im Rahmen der schweizerischen Gesetzgebung findet sich kein Artikel, welcher DRM ausdrücklich erwähnt. Dies ist durchaus einleuchtend, schliesslich gab es zu dem Zeitpunkt, als diese Gesetze erlassen wurden, noch keine technischen Schutzmassnahmen. Zurzeit sind die Gesetze zum Urheberrecht [1] sowie des Datenschutzes [2] sowie die dazugehörigen Verordnungen für DRM massgebend. Allerdings befindet sich das Urheberrecht momentan in Vernehmlassung, um den veränderten Bedürfnissen nachzukommen. DRM stellt eine technische Schutzmassnahme dar. Technische Schutzmassnahmen werden vom Gesetzgeber nur als eine Institution zum Schutz von Werken betrachtet. Weitere Institutionen des Schutzes sind Patente, der Urheberrechtsschutz, Schutz vor Unlauterem Wettbewerb, Schutz von Geschäftsgeheimnissen (hier vermutlich nur schlecht anwendbar), Schutz von Fabrikations- und Handelsmarken sowie vertragliche Schutzvorkehrungen.Dabei besonders interessant ist das Urheberrechtsgesetz:

#### Urheberrechtsgesetz

Der Urheber hat in der Schweiz das ausschliessliche Recht zu bestimmen, ob, wann, wie und unter welcher Urheberbezeichnung das eigene Werk erstmals veröffentlicht werden soll (Art. 9 Abs. 2 URG). Nach Art. 10 des URG hat der Urheber zudem das ausschliessliche

Recht, zu bestimmen, ob, wann und wie das Werk verwendet wird. Durch diese zwei genannten Artikel wird DRM implizit erlaubt - es ist dem Urheber des Werkes erlaubt, das Werk mittels DRM zu schützen. Allerdings macht das Urheberrecht auch Einschränkungen, wie weit die Rechte des Urhebers gehen. Diese Schranken des Urheberrechts werden im 5. Kapitel des Urheberrechtsgesetzes definiert:

- Art. 19 URG: Verwendung zum Eigengebrauch
  Veröffentlichte Werke dürfen zum Eigengebrauch verwendet werden (Abs. 1). Als Eigengebrauch gilt jede Werkverwendung im persönlichen Bereich und im Kreis von Personen, die unter sich eng verbunden sind, wie etwa Freunde oder Verwandte (Lit. a). Ebenfalls als Eigengebrauch zählt die Werkverwendung der Lehrperson für den Unterricht in der Klasse (Lit. b). Dazu zählt auch das Vervielfältigen von Werkexemplaren in Betrieben, öffentlichen Verwaltungen, Instituten, Kommissionen und ähnlichen Einrichtungen für die interne Information oder Dokumentation (Lit. c). Das heisst, ein DRM System muss diese Möglichkeit offen halten, schliesslich ist dies ein Recht, das vom Urheberrechtsgesetz ausdrücklich gewährt wird. Zu Artikel 19 stellt sich jedoch die Frage, ob Absatz 4 zur Anwendung kommt oder nicht. Absatz 4 lautet: 'Dieser Artikel findet keine Anwendung auf Computerprogramme.' Digitale Filme oder Musikstücke als Computerprogramme zu bezeichnen, dürfte jedoch sehr weit gefasst sein. Mit dieser Einschränkung wollte der Gesetzgeber schliesslich Software vor dem Weitergeben schützen.

- Art. 23 URG: Zwangslizenz zur Herstellung von Tonträgern
  Sehr interessant ist auch der Zwang des Gesetzgebers, dass auf Tonträger aufgenommener Werke der Musik mit oder ohne Text, welche mit der Zustimmung des Urhebers oder Urheberin angeboten, veräussert oder sonst wie verbreitet, an alle Hersteller und Herstellerinnen von Tonträgern mit einer gewerblichen Niederlassung im Inland vom Inhaber oder von der Inhaberin des Urheberrechts gegen Entgelt zwangslizenziert werden müssen (Abs. 1). Allerdings dürfte dieser Artikel kaum zum Tragen kommen, schliesslich sind digitale Musikstücke in der Regel eben nicht auf Tonträger aufgenommen. Hier stellt sich die Definitionsfrage eines Tonträgers.

- Art. 24 URG: Archivierungs- und Sicherungsexemplare
  Dieser Artikel erlaubt ausdrücklich, von Werken eine Kopie anzufertigen. Die Kopie muss dabei in einem der Allgemeinheit nicht zugänglichen Archiv aufbewahrt werden und ausdrücklich als Archivexemplar gekennzeichnet sein (Abs. 1). Diese Erlaubnis erstreckt sich auch auf Computerprogramme und kann nicht vertraglich wegbedungen werden (Abs. 2). Ein DRM System muss also diese Möglichkeit offen halten. Dies kann allerdings ein Problem darstellen - das System muss also zwischen einer erlaubten Archivkopie und einer nicht erlaubten Raubkopie unterscheiden können.

Besonders dieser letzte Punkt (Artikel 24 URG) dürfte ein grosses Hindernis sein, da Kopien erlaubt sein müssen, obwohl gerade dies vom Urheber ja mit dem DRM verhindert werden soll.

**Datenschutzgesetz**

Der Anbieter eines Werkes erhält mit DRM sehr viele Daten über den Kunden. Es ist durchaus realistisch, dass der Anbieter aufzeichnen kann, wer welches Werk wann wo abspielt. Dies stellt aus rechtlicher Sicht zunächst kein Problem dar – die Daten werden durchaus rechtmässig beschafft (Art. 4 Abs. 1 DSG). Diese beschafften Daten müssen jedoch zweckmässig gesichert werden, dies durch technische und organisatorische Massnahmen (Art. 7 Abs. 1 DSG). Das Datenschutzgesetz gewährt dem Kunden gegenüber dem Inhaber einer Datensammlung ein Auskunftsrecht (Art. 8 DSG), auf welches auch nicht im Voraus verzichtet werden kann (Art. 8 Abs. 6 DSG). Dieses Auskunftsrecht wird jedoch relativ selten genutzt und sollte den Anbieter auch nicht vor all zu grosse Probleme stellen. Problematischer ist jedoch die Tatsache, dass bei der Verwendung von DRM durchaus auch besonders schützenswerte Daten nach Artikel 3 Lit. c anfallen können, etwa Daten über die religiösen, weltanschaulichen, politischen oder gewerkschaftlichen Ansichten (Ziffer 1) oder die Intimsphäre oder die Rassenzugehörigkeit (Ziffer 2). Man stelle sich etwa vor, dass der User erotische Filme betrachte oder Dokumentarfilme zu bestimmten Krankheiten. Der Anbieter wüsste dann unter Umständen um die sexuellen Vorlieben des Anwenders oder um seinen Gesundheitszustand. Um diese besonders schützenswerten Daten bearbeiten zu können, muss der Inhaber der Datensammlung entweder die Zustimmung der betroffenen Personen erhalten oder die Datensammlung in das Register der Datensammlungen des Eidgenössischen Datenschutzbeauftragten eintragen lassen (Art. 11 DSG).

**Vernehmlassung der Änderung des Urheberrechtsgesetzes**

Zum Zeitpunkt des Verfassens dieses Dokuments befindet sich das Urheberrechtsgesetz in der Vernehmlassung[5][6]. Dies wurde nötig, um die Standards der Abkommen der Weltorganisation für geistiges Eigentum (WIPO) von 1996 sowie diejenigen der Europäischen Gemeinschaft zu berücksichtigen. Durch diese Umsetzung kann das vom Bundesrat unterzeichnete WIPO-Abkommen betreffend das Urheberrecht und die verwandten Schutzrechte ratifiziert werden. Das WIPO Copyright Treaty (WCT) und das WIPO Performances and Phonograms Treaty (WPPT) regeln den Schutz der Urheber, der Musikinterpreten sowie der Tonträgerhersteller in Bezug auf die grenzüberschreitenden Kommunikationstechnologien wie das Internet. Die deshalb auch als 'Internet-Abkommen' bezeichneten Verträge sind am 6. März (WCT) und am 20. Mai (WPPT) 2002 in Kraft getreten und haben die dazu notwendige Anzahl von 30 Ratifizierungen bzw. Beitritten inzwischen weit überschritten. Der Gesetzesentwurf enthält verschiedene neue Schutzausnahmen, die auch Werkverwendungen im digitalen Umfeld betreffen. Anderseits sieht er vor, dass die Inanspruchnahme der Schutzausnahme gegenüber dem Schutz technischer Massnahmen grundsätzlich Vorrang hat. Gleichzeitig wurden einige Vorstösse des Parlaments 'abgehandelt', die allerdings eher auf Details abzielten (Motion Christen (99.3557 Urheberrechtsentschädigung auf Subventionen) für eine bessere Berücksichtigung der Interessen der Nutzer bei der Aufstellung der Tarife durch die Verwertungsgesellschaften; Motion Weigelt (00.3127 Produzenten-Urheberrecht) für eine Verbesserung des Schutzes des Produzenten von urheberrechtlich geschützten Werken; Motion Aeppli Wartmann (01.3401 Folgerecht): Berücksichtigung Richtlinie 2001/84/EG vom 27. September 2001 über das

Folgerecht des Urhebers des Originals eines Kunstwerks). Die Motion Christen wurde dabei nicht ins Gesetz aufgenommen, da dies auf dem Verordnungsweg besser abgehandelt werden kann. Die Motion Weigelt wurde nicht umgesetzt, da sich trotz interessanter Kompromissvorschläge in der beratenden Kommission nur auf mangelnde Unterstützung von Produzentenseite her traf. Die Motion Aeppli Wartmann wurde ebenfalls nicht umgesetzt, da kein Kompromiss gefunden werden konnte (den Künstlerkreisen gingen die Kompromisse zu wenig weit, während der Kunsthandel die Kompromisse ablehnte). Ein weiterer Vorstoss war die Motion Thanei (04.3163 Gerätevergütung), welche eine Ergänzung des Vergütungssystems für das Vervielfältigen von Werken zum Eigengebrauch vorschlägt. Diese Motion wurde in die Revision aufgenommen (Art. 20a).

Das Urheberrecht wird wie folgt geändert (es wird hier nur auf die für dieses Dokument relevanten Änderungen eingegangen):

- Art. 19 Abs. 2 URG: Neu wird Bibliotheken ausdrücklich das Recht zugestanden, für ihre Benutzer eine Kopie eines Urheberrechtlich geschützten Dokuments herzustellen
- Art. 24c (neu) URG: Dieser neu geschaffene Artikel erlaubt ausdrücklich die Vervielfältigung von Werken für Menschen mit einer Behinderung, soweit die sinnliche Wahrnehmung der bereits veröffentlichte Form des Werkes für diese Personen nicht möglich oder erheblich erschwert ist.
- Art. 33 und 36 URG: Diese beiden Artikel wurden so erweitert, dass die Digitale Verbreitung nun auch abgedeckt ist ('oder sonst wie zu verbreiten').
- Titel 3a (neu) URG: Neu wird ein Titel 3a 'Schutz von technischen Massnahmen und von Informationen für die Wahrnehmung von Rechten' eingefügt. Dieser Titel regelt den Umgang mit technischen Schutzmassnahmen, ist also genau deser Titel, welcher DRM regelt. Die WIPO-Abkommen haben mit den sich aus Artikel 11 WCT und Artikel 18 WPPT ergebenden Verpflichtungen den Grundstein zu diesem neuartigen Schutz gelegt, der hier umgesetzt wird. Im Folgenden werden wir die einzelnen Artikel dieses neuen Titels kurz beleuchten:
  - Art. 39a (neu) URG: Schutz technischer Massnahmen
    * Absatz 1 verbietet das Umgehen technischer Schutzmassnahmen. Dies allerdings nur bis zum Ablaufen der gesetzlichen Schutzfrist. Es ist also verboten, DRM-Systeme zu umgehen - ausser die Schutzfrist des Werkes ist abgelaufen. In Übereinstimmung mit den Vorgaben der WIPO-Abkommen werden jedoch technische Massnahmen nicht generell geschützt, sondern nur insofern, als sie sich auf urheberrechtlich geschützte Werke oder Leistungen beziehen. Für technische Massnahmen, die gemeinfreie Werke und Leistungen bzw. ungeschützte Inhalte betreffen, besteht also kein Umgehungsverbot.
    * Absatz 3 verbietet das Herstellen, Einführen, Anbieten, Veräussern oder das sonstige Verbreiten, Vermieten, zum Gebrauch Überlassen, die Werbung für und der Besitz zu Erwerbszwecken von Vorrichtungen, Erzeugnissen oder Bestandteilen sowie die Erbringung von Dienstleistungen, die Gegenstand einer Verkaufsförderung, Werbung oder Vermarktung mit dem

Ziel der Umgehung technischer Massnahmen sind; abgesehen von der Umgehung technischer Massnahmen nur einen begrenzten wirtschaftlichen Zweck oder Nutzen haben; oder hauptsächlich entworfen, hergestellt, angepasst oder erbracht werden, um die Umgehung technischer Massnahmen zu ermöglichen oder zu erleichtern. Faktisch wird jegliche Möglichkeit zum Umgehen eines DRM verboten.

  - Absatz 4 macht jedoch eine Einschränkung zu dem Verbot: Danach ist ein Eingriff in die technischen Massnahmen zwar grundsätzlich verboten, aber seine Verletzung kann weder zivil- noch strafrechtlich geahndet werden, wenn der Eingriff ausschliesslich dem Zweck gedient hat, eine gesetzlich erlaubte Verwendung des Schutzobjekts vorzunehmen.

- Artikel 39b (neu) zwingt den Anbieter eines Werkes mit einer technischen Schutzmassnahme, deutlich erkennbare Angaben über die Eigenschaften der Massnahmen und die Identifizierung seiner Person machen. Durch diese Bestimmung soll Transparenz geschaffen werden und DRM geschützte Werke klar ersichtlich sein. Ausserdem hat der Nutzer gemässe Absatz 2 das Recht dazu, vom Anwender der technischen Schutzmassnahme Zugang zum DRM geschützten Werk zu erhalten, wenn er dazu befugt ist - deshalb auch die zwingende Identifizierung des Anwenders. Sollte der Anwender der technischen Schutzmassnahmen diesen Anforderungen nicht nachkommen, besteht nach Absatz 3 kein Anspruch mehr auf den Schutz der technischen Massnahmen. Zudem kann der Anwender mit einer Busse bis 20'000 Franken bestraft werden, wer diese Kennzeichnung vorsätzlich weglässt (Art. 70a (neu) URG).

**Konsequenzen aus der Revision des URG**

Die Änderungen des Urheberrechtsgesetzes bringen für den Bürger hauptsächlich neue Verbote: Es wird nicht nur verboten, Schutzmechanismen zu umgehen, sondern auch gleich alle Mittel dazu. Bei diesen wird sogar jede erdenkliche Möglichkeit verboten, wie die Menschen an sie kommen könnten (Herstellen, Einführen, Anbieten, Verkaufen, sonst wie Verbreiten, Vermieten, zum Gebrauch überlassen). Auch die Werbung für Mittel zur Umgehung von Schutzmassnamen wird verboten.

### 6.2.2 Rechtliche Situation in der EU

Massgebend in der EU ist die EUCD (Direktive 2001/29/EC), welche das europäische Gegenstück zum Digital Millennium Copyright Act (DMCA) in den USA darstellt (siehe nächster Abschnitt). Artikel 6 dieser Direktive verbietet dabei das Umgehen von effektiven technischen Schutzmassnahmen. Als technische Massnahme gilt dabei jegliche Technologie, Gerätschaft oder Komponente, welche dazu ausgelegt wurde, den Zugriff, welcher nicht vom Rechteinhaber erlaubt ist, zu beschränken oder verbieten. Als effektiv gilt die Massnahme bereits, wenn der Zugriff durch den Inhaber der Rechte durch Zugriffskontrolle oder einen Schutzprozess, etwa Verschlüsselung, Scrambling oder eine andere Art der Transformation des Werkes kontrolliert wird (Absatz 3). In Absatz 2 wird die Herstellung,

der Import, die Distribution, der Verkauf, das Vermieten, Werbung für den Verkauf oder die Vermietung oder der Besitz für kommerzielle Zwecke von Mitteln zur Umgehung von Schutzmassnahmen ausdrücklich verboten.

### 6.2.3 Rechtliche Situation in den USA

Die Rechtliche Situation in den USA ist geprägt durch den Digital Millennium Copyright Act (DMCA) von 1998. Dieses Gesetz geht noch weiter als die Anforderungen der WIPO. Hier ist der §1201, welcher das Umgehen von Technischen Mitteln zur effektiven Zugangskontrolle verbietet. Ebenfalls verboten wird in Abschnitt (b) die Herstellung, der Import, das öffentliche Anbieten, zur Verfügung stellen oder anderweitig verfügbar machen von Mitteln zur Umgehung von Schutzmassnahmen. Ausgenommen werden hier Nonprofit-Bibliotheken, Archive und Bildungseinrichtungen, jedoch nur um festzustellen, ob eine Kopie des Werkes erstanden werden soll. Ebenfalls Ausgenommen sind Strafverfolgungsbehörden, Geheimdienste und andere staatliche Aktivitäten. Reverse Engineering ist erlaubt, um die Schnittstellen zu einer Software zu bestimmen, soweit diese nicht bekannt sind. Des Weiteren ist die 'Verschlüsselungs-Forschung' erlaubt, das heisst, das Suchen von Fehlern und Verwundbarkeiten von Verschlüsselungsmethoden. Dieser heikle Punkt wird sehr genau definiert, da diese Ausnahme sonst zu einer Carte Blanche für Cracker ausgelegt werden könnte. Um sich nicht strafbar zu machen, muss die Person das verschlüsselte Werk legal erworben haben; das Umgehen der Schutzmassnahme ist notwendig für die Forschung; die Person muss einen aufrichtigen Versuch gemacht haben, eine Erlaubnis zu erhalten; die Forschung darf nicht zum Ziel der Piraterie geschehen; die Person muss effektiv in diesem Feld Studien betreiben oder angestellt sein oder ausreichend ausgebildet und erfahren sein; und der Inhaber der Rechte ist über die Ergebnisse der Forschung innert angebrachter Frist zu informieren.

### 6.2.4 Rechtliche Situation weltweit

International massgebend sind die WIPO-Abkommen betreffend das Urheberrecht und die verwandten Schutzrechte. Das WIPO Copyright Treaty (WCT) und das WIPO Performances and Phonograms Treaty (WPPT) regeln den Schutz der Urheber, der Musikinterpreten sowie der Tonträgerhersteller in Bezug auf die grenzüberschreitenden Kommunikationstechnologien wie das Internet und werden deshalb auch als 'Internet-Abkommen' bezeichnet. Der WCT ist am 6. März 2002 und der WPPT am 20. Mai 2002 in Kraft getreten, da beide die dazu notwendige Anzahl von 30 Ratifizierungen bzw. Beitritten inzwischen weit überschritten haben.

Alle führenden Industriestaaten haben die beiden Abkommen unterzeichnet und bereiten ihre Ratifikation vor.

## 6.3 Anforderungskatalog an ein mobiles DRM System

Aus den rechtlichen Verhältnissen und den Bedürfnissen der Industrie lässt sich nun ein Anforderungskatalog an ein (mobile) DRM-System erstellen. Viele Konzessionen sind von Seiten der Industrie freilich nicht zu machen; die Frage, ob Privatkopien überhaupt zulässig sind, ist derzeit Gegenstand von politischen Diskussionen.

Zu den offensichtlichen Anforderungen an ein Mobile-DRM-System gehört die Notwendigkeit einer effektiven Verschlüsselung für die geschützten Inhalte um Unbefugten den Zugriff zu verwehren. Dies folgt aus der Erkenntnis heraus, dass sich das Vervielfältigen von digitalen Daten tatsächlich kaum verhindern lässt.

Des weiteren muss eine Trennung von eigentlichem geschütztem Inhalt und der entsprechenden, dazugehörigen Rechteinformation vollzogen werden um beispielsweise einfache neue Distributionswege zu ermöglichen, was wiederum eine eindeutige Identifizierung eines geschützten Inhaltes erforderlich macht. Da nicht mit permanenter Zugriffsmöglichkeit auf die Rechteinformationen beispielsweise auf einem zentralen Server gerechnet werden kann, muss die Rechteinformation zu einem geschützten Inhalt auch auf der Instanz, auf dem sich der geschützte Inhalt befindet, vorhanden bzw. veränderbar sein. Die Rechteinformation muss ausserdem an ein oder mehrere Endgeräte bindbar sein, um vervielfältigung zu verhindern. Ausserdem muss sichergestellt werden, dass das entsprechende Endgerät die entsprechenden Rechteinformationen auch korrekt und genau umsetzt.

Die Anforderungen an ein DRM-System sollen im Folgenden anhand des Anforderungskataloges der Open Mobile Alliance (OMA) für Mobile DRM in der Version 2 [14] diskutiert werden.

### 6.3.1 Der OMA DRM Anforderungskatalog

Der OMA-Katalog entwickelt sich aus Szenarios, welche das Soll-Verhalten des Systemes anhand von praktischen Fällen beschreiben. Ein solches Szenario im Katalog ist beispielsweise folgendes Beispiel, was das Systemverhalten im Falle einer Beschädigung des DRM-Agentsystems beschreibt:

> Jo drops her mobile Device resulting in a catastrophic failure, she calls her Network Service Provider who replaces the Device (under her insurance agreement). The embedded portable smartcard Device carries her identity in a secure way. The smartcard has not been damaged and she is able to insert it in the replacement Device and use this as an authenticated identity which allows her to download the DRM Content and Rights Object previously purchased from the Content Providers.

Das aktuelle OMA-DRM Anforderungsdokument beinhaltet mehr als 20 solche Szenarios, welche die wichtigsten Situationen abdecken sollen. Aus diesen Szenarios werden schliesslich neben technischen Entwicklungsanforderungen die Marktanforderungen abgeleitet, welche nachfolgend aus dem Englischen übersetzt gezeigt werden:

1. 'Es soll möglich sein, geschützen Inhalt präzise zu identifizieren so dass diesem ein Rechteobjekt eindeutig zugewiesen werden kann.'

   Dies ist nowendig, da das Rechteobjekt nur für einen spezifischen, entsprechenden Inhalt gültig sein darf; also muss dieser auch identifiziert werden können.

2. 'Es soll für einen Rechteherausgeber möglich sein, Rechteobjekte an Geräte zu schicken.'

   Dieser Punkt unterstreicht, dass die Rechte von einer 'externen' Stelle verwaltet und verteilt werden; er impliziert auch, dass die Kontrolle über die Rechteinformationen von einer entfernten Instanz verwaltet werden.

3. 'Die Berechtigungen in einem Rechteobjekt sollen von einem OMA-konformem Gerät durchgesetzt werden.'

   Dies ist eine kritische Anforderung für mobiles DRM; es ist schliesslich nutzlos, wenn der Inhalt geschützt ist, die Rechte auf dem Endgerät aber nicht durchgesetzt werden können oder gar Tür und Tor zu unerlaubten Kopien geöffnet werden. Man spricht in diesem Zusammenhang auch von einem Trusted Device (Vertrauenswürdiges Gerät), was hohe Anforderungen an die Implementierung stellt.

4. 'Es soll für einen DRM Agent nicht möglich sein geschützen Inhalt zu verwenden, ohne dass diesem ein entprechendes Rechteobjekt zugeordnet wurde und der Agent dieses besitzt.'

   DRM geschützte Inhalte können zwar auf das Gerät übertragen werden, sind aber ohne Rechteobjekt nutzlos.

5. 'Rechteobjekt und geschützter Inhalt sollen physisch, nicht aber logisch getrennt werden können.'

   Rechteinformation und Inhalte sind eng aneinander gekoppelt, im Extremfall ist ein geschützter Inhalt ohne entsprechendes Rechteobjekt nicht mehr nutzbar und somit wertlos.

6. 'Es soll möglich sein, Rechteobjekt und geschützen Inhalt auf demselben oder auf unterschiedlichen Transportwegen auszuliefern. Die Auslieferung soll mit beliebigen Transportmechanismen möglich sein.'

   Dies ist eine Voraussetzung, um die grossflächige Verteilung der eigentlichen Inhalte beispielsweise über Superdistribution zu ermöglichen, während die Rechte dazu zentral von einer autorisierten Instanz verwaltet werden können.

7. 'Geschützer Inhalt kann Medienobjekte beliebigen Inhaltstyps beinhalten.'

   Damit soll sichergestellt werden, dass unterschiedliche Endapplikationen wie Musik, Video oder Dokumentverteilung/verarbeitung von der DRM Archtitektur ermöglicht werden und sich das System nicht auf einzelne feste Formate beschränkt (wie

dies derzeit beipsielsweise bei iTunes der Fall ist). Damit soll wohl auch dem Drang der einzelnen Hersteller entgegengewirkt werden, die DRM-Architektur an ihre eigenen, proprietären Inhaltsformate binden zu wollen. Im Hinblick auf die Fähigkeit einzelner Formate, die Marktstellung eines einzelnen Herstellers zu zementieren (beispielsweise das .doc Format von Microsoft Word) ist eine formatunabhängige Architektur absolut zwingend, sofern ein offener, wettbewerb-ermöglichender Standard etabliert werden soll.

8. 'Eine Überprüfung auf dem Endgerät auf Abspieltauglichkeit für ein Objekt soll möglich sein, bevor ein entprechendes Rechteobjekt für das entprechende Objekt mit geschütztem Inhalt angefordert wird.'

   Damit soll gewährleistet werden, dass vom Benutzer keine Inhalte erworben werden können, welche auf seinem Gerät gar nicht genutzt werden können; somit können im Voraus Frustration, Reklamationen, Beschwerden oder gar Klagen vermieden werden.

9. 'Es soll für den Rechteherausgeber möglich sein, die Abspielfähigkeit eines Gerätes für ein bestimmtes Objekt mit geschützem Inhalt festzustellen, bevor das entprechende Rechteobjekt an das Gerät geschickt wird.'

   Die Überprüfung soll nicht nur wie oben erwähnt auf Benutzerseite möglich sein, sondern auch vom Anbieter bzw. Rechteherausgeber aus; dies erhöht die Sicherheit zusätzlich (Im Falle des Versagens einer Prüfseite), im weiteren treffen dieselben Argumente zu, die beim vorherigen Punkt erwähnt werden.

10. 'Berechtigungen innerhalb des Rechteobjektes sollen die folgenden Eigenschaften enhalten. Alle Berechtigungen sollen ausdrücklich ausgewiesen werden:

    (a) Es soll möglich sein, Berechtigungen für die folgenden Abspielmethoden zu spezifizieren:
        - Abspielen
        - Ausführen
        - Anzeigen
        - Drucken

    (b) Folgende Nutzungsbeschränkungen sollen möglich sein:
        - Zeit / Datumsbasiert
        - Limitiert in der Anzahl der Benutzungen'

    Diese Anforderungen erlauben eine feingranulare Aufteilung und Verwaltung der Rechteinformation über ein Objekt. Dem Kunden können auf diese Weise sehr spezifische, eingeschränkte Angebote gemacht werden, was sowohl aus Kundensicht (niedrigere Preise) wie auch für den Anbieter (neue Geschäftsmöglichkeiten, beispielsweise Inhalte zum Einmalgebrauch) interessant sein kann.

11. 'Berechtigungen innerhalb der Rechteobjekte sollen folgende Eigenschaften aufweisen:

    (a) Es soll möglich sein, sowohl Rechteobjekte wie auch geschützte Inhalte von einem Gerät auf ein anderes zu exportieren

(b) Folgende Nutzungsbeschränkungen sollen möglich sein:
- Basiert auf Zeitmessung (d.h. Ein Gerät kann geschützten Inhalt solange abspielen, als die verwendete Nutzungszeit kleiner als die spezifierte ist)
- Basiert auf der Benutzeridentität (d.h. Ein Gerät kann ausschliesslich geschützten Inhalt wiedergeben wenn es von einem definierten Benutzer verwendet wird)'.

Einerseits soll hier die Wichtigkeit der Transferierbarkeit von Rechten und Inhalten betont werden; dies erlaubt dem Kunden, seine bestehenden Inhalte auch auf einem neuerworbenen Gerät abzuspielen. Im zweiten Teil dieses Punktes werden weitere Einschränkungsmöglichkeiten spezifiziert; einerseits soll eine zeitliche Limiterung der möglichen Abspieldauer möglich sein, andererseits soll eine Bindung an einen spezifischen Benutzer (nicht zu verwechseln mit Geräten) ermöglicht werden. Auch diese Anforderungen sollen neue Angebote ermöglichen, welche noch stärker personalisiert werden können.

12. 'Es soll möglich sein, sowohl DRM Inhalte wie auch zustandlose Rechteobjekte von einem Gerät auf ein Backup-Medium zu sichern.'

    Mit dieser Anforderung soll der Tatsache Rechnung getragen werden, dass es sich um digitale Daten handelt, welche unter Umständen verloren gehen können. Ausserdem wird eine Backup-Möglichkeit von vielen Ländern gesetzlich vorgeschrieben/ausdrücklich erlaubt; Diesbezügliche Konformität ist also kritisch für den internationalen Erfolg eines DRM Systemes.

13. 'Es soll für den Rechteherausgeber möglich sein, ein Gerät zuverlässig zu identifizieren, um diesem Gerät Rechteobjekte zukommen zu lassen oder zu verweigern.'

    Dank einer eindeutigen Identifizierung können 'schwarze Schafe', also Geräte, welche aus bestimmten Gründen vom System ausgeschlossen werden sollen, erkannt und blockiert werden. Dies könnte zum Beispiel Geräte betreffen, deren Trusted Device Core nicht mehr als sicher betrachtet werden kann, wenn der Besitzer eigenhändig Änderungen an der Software vorgenommen hat. Es wird so theoretisch aber auch möglich, alte Gerätegenerationen, bei welchen sich beispielsweise ein Herstellungsfehler im Trusted Device Programmcode herausgestellt hat, zukünftig vom Geschäft auszuschliessen.

14. 'Für Rechteherausgeber soll die Möglichkeit bestehen, Rechteobjekte auf eine bestimmte Gruppe von Geräten zu beschränken, so dass das Rechtsobjekt nur von einer beabsichtigten Gruppe besessen werden kann.'

    Diese Anforderung kann wiederum für die Etablierung neuer Geschäftsmodelle verwendet werden, beispielsweise für Promotionen von spezifischen Gerätetypen. Ausserdem kann so auch verhindert werden, dass hochwertige Inhalte auf 'minderwertigen' Geräten abgespielt werden können.

15. 'Geräte sollen Rechteobjekte an andere Geräte senden können (Das empfangende Gerät kann das Rechteobjekt jedoch nur verarbeiten, wenn der Rechteherausgeber dies zulässt).'

    So kann die direkte Weitergabe von Inhalten im Peer-to-Peer-Stil (p2p) erlaubt werden, falls dies vom Rechteinhaber erwünscht ist.

16. 'DRM Inhalte und Rechteobjekte sollen gleichzeitig oder versetzt ausgeliefert und in beliebiger Reihenfolge empfangen werden können.'

    Dieser Punkt zielt in dieselbe Richtung wie Punkt 6; es wird lediglich zusätzlich betont, dass auch die Reihenfolge der Objekttransfers irrelevant sein soll.

17. 'Ein Gerät soll die Integrität von Superdistributionsinhalten feststellen können.'

    Durch einen Integritätscheck können Übertragungsfehler sowie absichtliche Veränderungen an einem Originalpaket erkannt und der entsprechende Inhalt blockiert werden. Dies dient wohl der Verhinderung des Weiterkopierens von bereits entschlüsselten Inhalten.

18. 'Mehrere Inhalte sollen in einem einzelnen Paket verteilt und vom User heruntergeladen werden können, während für jedes Objekt in diesem Paket unterschiedliche Rechteeigenschaften definiert werden können.'

    Damit kann ein Bundling von verschiedenen Inhalten angeboten werden; ausserdem ist so der Vertrieb von Komplettpaketen, wie beispielsweise einer ganzen CD (im Gegensatz zu einzelnen Liedern) denkbar. Ein weiteres Szenario wäre der Vertrieb eines Demo-Inhaltes, beispielsweise eines Musiktitels, von dem kostenlos nur die ersten 30 Sekunden angehört werden können, zusammen im Paket mit dem kompletten Song, für dessen Nutzung jedoch dann ein Rechteobjekt erworben werden muss.

19. 'Ein Gerät soll DRM-Inhalte abspielen können, die von einem Backup wiederhergestellt wurden.'

    Die Restore-Möglichkeit von einem Backup wird somit ebenfalls gewährleistet; schliesslich macht ein Backup ohne Wiederherstellungsmöglichkeit wenig Sinn.

20. 'Ein Gerät soll geschützen Inhalt und verschlüsselte Rechteobjekte auf ein anderes Gerät kopieren können, welches nicht notwendigerweise über Netzwerkzugang verfügt (z.B. von einem Telefon auf einen tragbares Musikabspielgerät).'

    Die Palette an Zielgeräten wird somit erheblich verbreitert. Es können auch proprietäre Verbindungsmöglichkeiten zu den enstprechenden Endgeräten verwendet werden. Das DRM-Konzept erfordert es jedoch, dass auch diese Geräte mit entsprechender DRM-Software ausgerüstet sind, so dass die Inhalte verschlüsselt gespeichert werden können; andernfalls ist der kopierschutz nicht mehr gewährleistet. Bereits existierende, konventionelle Geräte wie z.B. MP3-Player sind also kaum kompatibel und somit nicht nutzbar in Verbindung mit DRM-Technologie, es sei denn, der Rechteanbieter erlaubt den Export (im Daten-Klartext) auf solche Geräte, ähnlich wie beispielsweise bei iTunes das Erstellen einer (ungeschützten) Audio CD erlaubt ist.

Neben den Marktanforderungen beinhaltet der OMA-Anforderungskatalog auch sehr umfangreiche Anforderungen an das Engineering, Sicherheit, Verrechnung, Rechte und andere Domänen. Diese hier ebenfalls näher zu beleuchten würde den Rahmen dieses Papers allerdings sprengen, weshalb auf eine genauere Betrachtung verzichtet wird.

Die Anforderungen bis und mit Aufzählungspunkt 10 gehörten auch schon zum Anforderungskatalog von OMA DRM 1.0, für welches heute erste Lösungen auf dem Markt

sind. Die weiteren Punkte wurden für den kommenden Standard OMA DRM 2.0 ergänzt. Entsprechende Produkte gibt es derzeit aber noch keine auf dem Markt.

Beim direkten Vergleich ist der Schwerpunkt des neuen Standards deutlich erkennbar: Die Mobilität von DRM-geschützten Inhalten zwischen verschiedenen Endgeräten gewann deutlich an Bedeutung; dies ist wohl nicht zuletzt auf die starke Verbreitung von 'simplen' tragbaren Abspielgeräten ohne Netzzugriff zurückzuführen, welche von den bisherigen Anforderungen nicht erfasst wurden.

## 6.4 Überblick und Diskussion von aktuellen Verfahren und Standards

### 6.4.1 Industriestandards und allgemeine Verfahren

Die Entwicklung von Mobile DRM wird primär von der Open Mobile Alliance (OMA) vorangetrieben. Die OMA ist ein Zusammenschluss von ungefähr 400 Unternehmen, unter anderem aus den Branchen IT, Netzbetreiber und Endgerätehersteller, deren erklärtes Ziel es ist, für Standards und Interoperabilität im mobilen DRM-Bereich zu sorgen. Das Resultat der Bemühungen sind die DRM Standards OMA DRM 1.0 und OMA DRM 2.0.

Die OMA-Standards sind derzeit jedoch noch nicht etabliert, von OMA DRM 1.0 gibt es lediglich eine Handvoll Implementierungen, von OMA DRM 2.0 noch gar keine. Überdies derzeit herrscht ein öffentlich ausgetragener Disput zwischen der GSM Association (GSMA) und der MPEG LA. Die GSMA, ein Zusammenschluss von Mobilfunkbetreibern ist nicht gewillt, den Lizenzforderungen der MPEG-LA für Patentlizenzen nachzukommen. Die MPEG-LA, verwaltet einen Pool von Patenten, welche für Implementierungen von OMA DRM (zumindest gemäss der MPEG-LA) notwendig sind. Die MPEG-LA hatte die Lizenzierungkonditionen bereits kürzlich gesenkt, um der GSMA entgegenzukommen.[8]

#### Eine Mobile-DRM-Architektur: OMA DRM Architecture

Anhand der Referenzarchitektur der Open Mobile Alliance [3] soll hier knapp erläutert werden, wie ein System für Mobiles DRM grundsätzlich aufgebaut ist.

Innerhalb der OMA-Architektur wird (wie auch in den meisten anderen Systemen) hauptsächlich zwischen folgenden Rollen unterschieden:

- Der **DRM Agent** repräsentiert eine sog. 'vertrauenswürdige' Instanz, welche für die Durchsetzung der Rechte und Einschränkungen von geschützten Inhalten ist. Konkret handelt es sich dabei eigntlich um die DRM-Software auf dem Endgerät. Vertrauenswürdig heisst hier, dass sich der Rechteherausgeber darauf verlässt, dass der Agent die durch ihn zugestandenen Rechte und Beschränkungen durchsetzt.

- Der **Inhalteherausgeber** stellt die Inhalte bereit, welche durch das DRM-System geschützt werden sollen.
- Der **Rechteherausgeber** ist diejenige Instanz, welche den DRM-Inhalten Rechte bzw. Beschränkungen zuweist und Rechteobjekte erstellt. Ein **Rechteobjekt** wiederum ist im wesentlichen ein XML-Dokument, welches die Rechteeigenschaften und Nutzungsmöglichkeiten eines bestimmten DRM-Inhaltes formuliert. Es handelt sich dabei um ein spezielles XML Dokument.
- Der **Benutzer** repräsentiert den (menschlichen) Endkunden, welcher auf DRM-geschützte Inhalte über einen DRM Agenten zugreift.

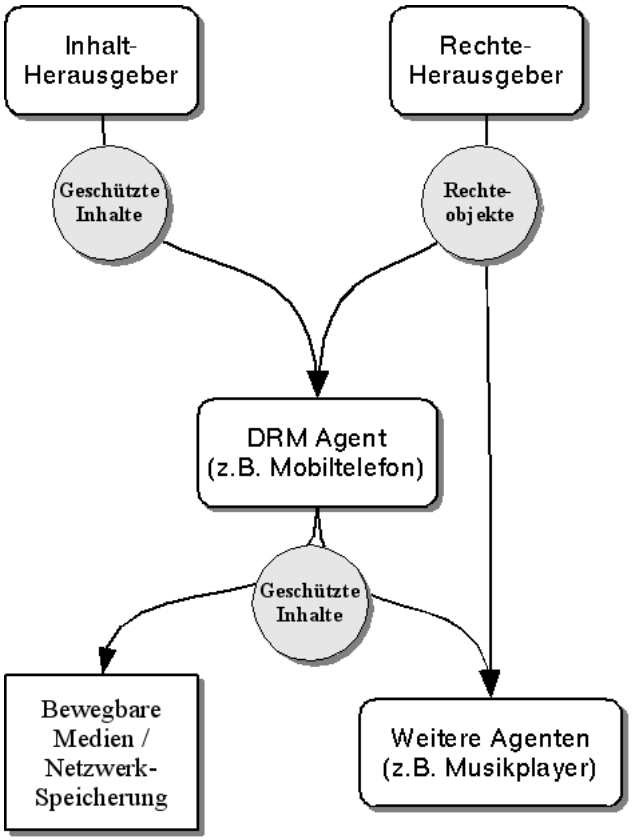


Abbildung 6.3: Die OMA Mobile DRM Architektur

Abbildung 6.3 verdeutlicht das Zusammenspiel zwischen den verschiedenen Akteuren in der OMA mobile DRM Architektur: Inhaltherausgeber und Rechteherausgeber sind getrennte Instanzen, welche Ihre Objekte an den DRM-Agenten sowohl gleichzeitig über den selben Kanal wie auch zeitlich versetzt über unterschiedliche Kanäle ausliefern können. Der DRM-Agent wiederum kann zwar die geschützten DRM-Inhalte an andere DRM-Agenten oder an Speichermedien (z.B. Memory Cards usw.) weiterverteilen oder kopieren. Das zum geschützten Inhalt zugehörige Rechteobjekt ist jedoch kryptografisch an den einzelnen DRM-Agenten gekoppelt und wird zum Verwenden der geschützten Inhalte benötigt, so dass jede weitere DRM-Agenteninstanz, welche von einem anderen DRM-Agenten einen geschützten Inhalt erhält, beim Rechteherausgeber ein neues Rechteobjekt anfordern muss. Optional kann der Rechteherausgeber ein Rechteobjekt nicht nur an einen

einzelnen Agenten binden, sondern auch an eine Gruppe von Agenten (eine sog. 'Domäne'). Somit lassen sich (zumindest in der Theorie) individuelle Angebote für ganze Geräte- bzw. Benutzergruppen realisieren.

### 6.4.2 Proprietäre Standards und Produkte

Neben den Anstrengungen der Industrie, wie sie z.B. die OMA verfolgt, einen relativ offenen Standard zu etablieren, gibt es diverse Firmen, welche eigene, proprietäre DRM-Systeme auf dem Markt anbieten. Erstaunlich ist dabei, dass ein Teil dieser Firmen auch Mitglieder der OMA sind, also gewissermassen eine Zweigleise-Strategie fahren.

Wie auch beispielsweise im Whitepaper von Berlecon [7] vermerkt ist die Trennung zwischen offenen und proprietären Standards schwierig zu vollziehen. In diesem Kontext definieren wir offene Standards als Standards, welche das Resultat gemeinsamer Anstrengungen grosser Teile der massgebenden (auch branchenübergreifenden) Industriemitglieder sind, währenddessen wir proprietäre Standards als Eigenentwicklungen einzelner Hersteller betrachten, zu deren Spezifikationen, Verfahren usw. fremde Marktteilnehmer keinen Zugang resp. Mitsprachemöglichkeit erhalten. Ein offener Standard bedeutet in diesem Zusammenhang keinesfalls, dass die Verfahren von allen Interessenten kostenlos genutzt werden können; Die unter dem Punkt 'Industriestandards und allgemeine Verfahren' erwähnten Dispute um Lizenzieungsfragen illustrieren das grosse Interesse an der patentrechtlichen Vermarktung dieser Verfahren.

#### FairPlay / iTunes von Apple

Seit längerem betreibt Apple mit iTunes einen Online Shop für Musik, seit neuestem auch in der Schweiz.[9] Wie zu Beginn dieses Dokumentes erwähnt, verwendet Apple mit 'FairPlay' ein eigenes DRM-System für den Schutz seiner angebotenen Inhalte. Apple nimmt im Vergleich zu den anderen hier vorgestellten Produkten insofern eine Sonderrolle ein, als dass FairPlay nicht zuletzt eingesetzt wird, um die Kunden an das ebenfalls von Apple hergestellte portable iPod Musikabspielgerät zu binden.[10], weshalb Kunden auch schon gegen Apple vor Gericht zogen.

Technisch gesehen ist Fairplay eine verhältnissmässig einfaches DRM-System. Die iTunes Clientsoftware verwendet einige Sytemdaten, um für den Computer einen eindeutigen Identifikationscode zu generieren. Dieser Hash wird an den iTunes-Server gesendet. Auf dem Server können bis zu drei Hashes zu einem Konto abgelegt werden. Der Server sendet dem Client anschliessend einen 'Account decryption key' zurück. Dieser Schlüssel wird in iTunes lokal gespeichert und mit dem Hash verschlüsselt, damit er nicht auf einen anderen Computer transferiert werden kann. Abbildung 6.4 soll diesen Vorgang verdeutlichen.

Wenn ein Musikstück nun in der iTunes-Software abgespielt wird, wird der decryption key mittels dem Hash-Wert des Computer wieder entschlüsselt. Anschliessend wird dieser decryption key dazu verwendet, das Musikstück zu entschlüsseln.

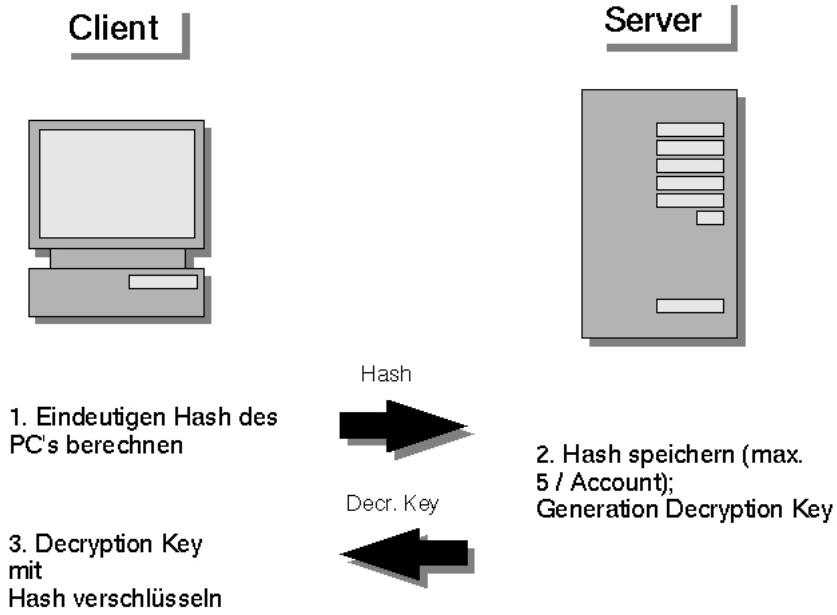


Abbildung 6.4: Diagramm zur Funktionsweise von Fairplay

Wenn nun ein Musikstück auf den iPod transferiert wird, wird der decryption key mittels dem Hashwert des Computers wieder entschlüsselt und anschliessend mit dem Hash-Wert des iPod's verschlüsselt. Dies gewährleistet, dass keine Musikstücke von iPod zu iPod kopiert werden können.

Um einen Computer zu deautorisieren, läuft der ursprüngliche Prozess in umgekehrter Reihenfolge ab: Der Hash-Wert wird an den Apple-Server geschickt, dort aus der Liste des Accounts gelöscht und der decryption key auf dem lokalen Rechner gelöscht. Dies führt zu einem Loch im System, welches dazu führt, dass beliebig viele Computer autorisiert werden können: Zuerst wird der Computer autorisiert, ein Backup vom verschlüsselten Decryptionkey gemacht, der Computer de-autorisiert, und der zuvor gespeicherte decryption key aus dem Backup wiederhergestellt. Nun denkt der Computer, er sei autorisiert, um Musikstücke abzuspielen, während der iTunes Music Store weiss, dass er es nicht ist und deshalb die Autorisierung von weiteren PCs erlaubt.[11]

Im Vergleich zum oben erläuterten Anforderungskatalog erfüllt iTunes die Anforderungen nur teilweise. Insbesondere sind folgende Anforderungen nicht erfüllt:

- Die geschützten Inhalte sind auf Musikstücke in einem festen Format limitiert
- Es werden keine Berechtigungsstufen unterschieden (Abspielen, Ausführen, Anzeigen, usw.)
- Nutzungsbeschränkungen nach Anzahl und Zeit sind nicht möglich
- Rechte sind nicht direkt von einem iPod auf einen anderen transferierbar
- Inhaltsverteilung per Superdistribution ist nicht vorgesehen

### Windows Media DRM von Microsoft

Der Software-Gigant Microsoft scheint betreffend DRM gleichzeitig verschiedene Strategien zu verfolgen. So ist Microsoft zwar einerseits in den Industriekonsortien zur Schaffung von offenen Standards vertreten, gleichzeitig hat der Softwaregigant aber auch eigene, proprietäre Lösungen im Angebot und unternimmt grosse Anstrengungen, diese im PC- und Mobilfunkmarkt voranzubringen. Konkretes DRM-Produkt stellt dabei Windows Media DRM (gegenwärtig Version 10) dar.

Erwartungsgemäss sind für die proprietäre Microsoft-technologie Detailspezifikationen für Windows Media DRM nur schwer zu finden; Microsoft verkauft unter dem Begriff Windows Media DRM verschiedene Lösungen, unserer Diskussion entspricht am ehesten 'Windows Media DRM 10 for Portable Devices'. Dieses scheint dem OMA-Modell sehr ähnlich zu sein, ein 'License Server' übernimmt die Rolle des Rechteherausgebers, und Distribution von Inhalten ist sowohl direkt via Internet als auch indirekt von Gerät zu Gerät möglich.[12] Auch die Nutzungsrestriktionen(beschränkte Abspieldauer,beschränkte Anzahl an Nutzungen) scheinen weitgehend vorhanden zu sein. Weiter als die anderen Produkte geht Windows Media DRM insofern, dass bereits bestimmte Geschäftsmechanismen (Abo-Systeme, Subskriptionsdienste) vorgesehen sind und von Microsoft wohl als separate Produkte vermarktet werden. Ausserdem versucht der Hersteller, die einzelnen Komponenten und Produkte möglichst miteinander zu verflechtem, z.b. die DRM-Lösung mit der entsprechenden Windows-Desktopsoftware für die Erstellung von geschützten Inhalten.

### Helix DRM 10 von Real

RealNeworks hat mit Helix DRM ebenfalls eine proprietäre DRM-Lösung auf dem Markt. Details zu den detaillierten Verfahren und Spezifikationen sind hier schwerer zu finden - aus offiziellen technischen Angaben [4] lässt aber auf ähnliche Funktionalität wie in der OMA 1.0 Spezifikation schliessen: so scheint das Helix DRM System im Wesentlichen aus folgenden Komponenten zu bestehen:

- Helix DRM Packager, um digitale Inhalte zu verschlüsseln und zur Distribution vorzubereiten (entspricht der OMA-Instanz 'Inhalteherausgeber')
- Helix DRM License Server, um Lizenzanfragen von Clients zu beantworten und diesen entsprechende Lizenzobjekte zukommen zu lassen bzw. zu entreissen. (entspricht der OMA-DRM-Instanz 'Rechteherausgeber')
- Helix DRM Client , um die geschützten Inhalte entsprechend den Lizenzinformationen abzuspielen (entspricht der OMA-DRM-Instanz 'DRM-Agent')

Wie in der OMA-Spezifikation wird auch hier die Trennung von (geschütztem) Inhalt und Lizenz- bzw. Rechteobjekt vollzogen ('The packaged media content and the associated business rules for unlocking and using that content are stored separately[...]').

Zusätzlich weist das System den 'Helix DRM Device Support' auf, welcher es u.a. ermöglichen soll, DRM-Objekte von sogenannten 'Primärgeräten' (beispielsweise PC mit

Internet-Anbindung) auf ebenfalls gesicherte Speichermedien (Sekundärgeräte, beispielsweise portabler Musikplayer) zu transferieren.

#### Adobe Adrobat / LifeCycle

Der Software-Hersteller Adobe hat ebenfalls eine DRM-Lösung im Angebot, die jedoch eher auf das vertrauliche Management von Dokumenten zielt als auf Nutzungsregelungen von Unterhaltungsmedien. Verglichen mit unserer Referenzarchitektur übernimmt der Adobe LiveCycle Policy Server gewissermassen die Rolle des Rechteherausgebers, während Acrobat (oder automatisierte Verfahren) als Tools für die Erstellung und Herausgabe der Inhalte (Inhaltherausgeber) gedacht sind. Der bekannte Acrobat Reader übernimmt schliesslich die Funktionalität des DRM-Agenten.

## 6.5 Kritik an (mobile) DRM

Viele Organisationen und bekannte Fachleute kritisieren die DRM-Pläne der Industrie und lehnen diese teilweise oder komplett ab. Die Kritik ist sehr vielseitig und betrifft verschiedenste Aspekte der DRM-Technologie.

Ein häufig genannter Kritikpunkt sind beispielsweise die durch DRM-Verfahren beschnittenen Rechte, Freiheiten und Möglichkeiten des Besitzers. Das erklärt auch, weshalb viele Kritiker 'DRM' mit 'Digital Restrictions Management' übersetzen, da es ihrer Meinung nach nicht um die Durchsetzung von Rechten des Kunden geht, sondern um deren Einschränkung. Es erstaunt deshalb auch nicht besonders, dass von Seiten der Befürworter freier Software, insbesondere den Verfechtern der GNU-Philosophie wie Richard Stallman, DRM kritisch begegnet wird; dies hauptsächlich, weil die künstlichen Einschränkungen von DRM im diametralen Gegensatz zur GNU-Lehre stehen, nach der Inhalte frei und für jedermann möglichst einfach zugänglich und nutzbar sein sollten. [13]

Ein weiterer Kritikpunkt betrifft die Sicherheit der Implementierungen von DRM Architekturen. Es steht dabei die Frage im Vordergrund, ob es überhaupt möglich ist, ein System von derartiger Komplexität grossflächig einzusetzen, ohne durch Fehler und Limitationen in den Implementierungen seine Integrität zu gefährden. So sind beispielsweise die meisten kryptografischen Methoden und deren Implementierungen nach wie vor Gegenstand von Kontroversen und Unsicherheiten (Siehe Fall Adobe Acrobat Verschlüsselung). Auch ist es trotz aller Bestrebungen möglich, die (entschlüsselten) Inhalte spätestens bei deren Ausgabe auf einen Bildschirm, Lautprecher usw. abzufangen und frei von jeglichen Schutzfunktionen weiterzuverwenden. Betrachtet man beispielsweise iTunes als derzeit wohl meistverbreitetstes DRM-System am Markt, scheinen sich gewisse Prophezeihungen zu erfüllen: Tatsächlich tauchten bereits verschiedene Hacks auf, welche es ermöglichen, die im iTunes Music Store erworbenen Inhalte (wider den Willen der Betreiber) ohne DRM-Schutz abzuspeichern. Ebenso lassen sich beispielsweise einmal im Sinne der Betreiber auf Audio-CD gebrannte Inhalte anschliessend problemlos wieder einlesen, codieren und im Internet ebenfalls ohne Schutz weiterverbreiten. Der Comic in Abbildung 6.5 illustriert diese Situation (Anmerkung: 'RIAA' steht für 'Recording Industry Association of

America', eine Vereinigung von Musikkonzernen, welche vehement gegen illegale Kopien vorgehen und entsprechend DRM und Kopierschutztechniken fördern).

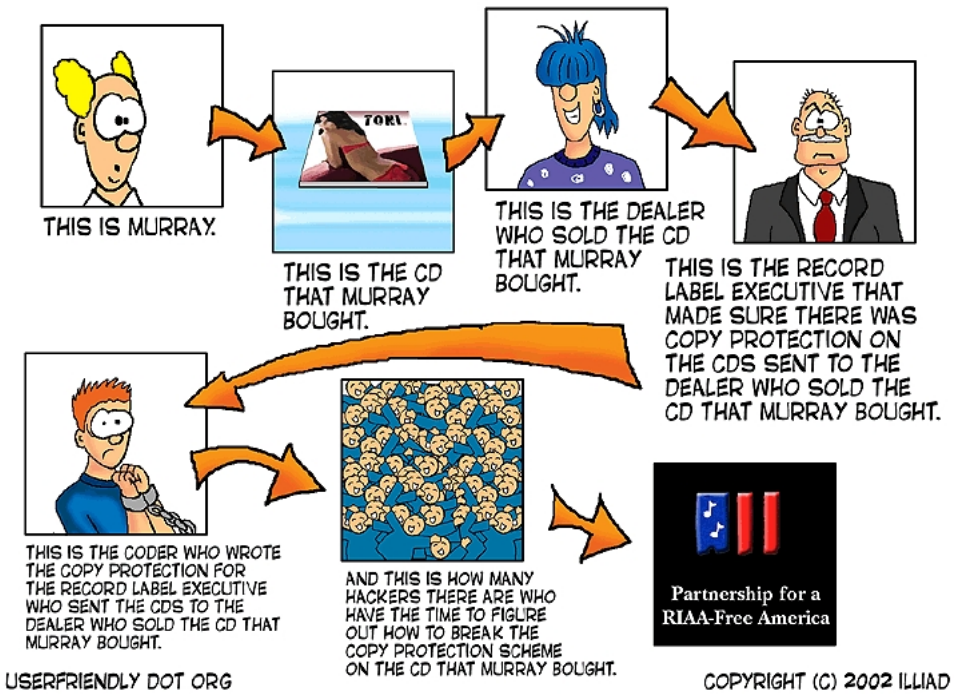


Abbildung 6.5: Comic 'Partnership for a RIAA-Free America'

Ebenfalls auf Widerstand stösst die Tatsache, dass mit mobilen DRM-Systemen effektiv nur noch eine Nutzungslizenz für einen Inhalt erworben wird und nicht mehr der Inhalt selbst. Viele Konsumenten möchten beispielsweise eine CD mit Musik besitzen (Im Sinne von Eigentum), und nicht nur über eine beschränkte Nutzungslizenz verfügen, welche die genauen Verwendungsmöglichkeiten des Produktes bis in die Details vorschreibt und möglicherweise erst noch nachträglich vom Herausgeber widerrufen oder verändert werden kann. Mit der Vorstellung, dass man es in Zukunft mit immer komplexeren 'Nutzungsverträgen' für Medien zu tun haben würde anstatt mit einem einem physischen Objekt, welches man sein Eigentum nennen darf, können sich viele Personen (verständlicherweise) nicht anfreunden.

Die Angst vor zunehmender Kommerzialisierung der bereits jetzt sehr intensiv vermarkteten künstlerischen Werke sorgt für zusätzliches Kopfzerbrechen vieler Konsumenten. Es deutet einiges darauf hin, dass die Industrie essentielle Kosteneinsparungen, welche beispielsweise beim Einsatz von DRM zum Medien-Direktvertrieb über Internet erreicht werden, nicht unbedingt an die Konsumenten weitergibt. Die Preise des iTunes-Musicstore Schweiz für ein aktuelles Album unterscheiden sich beispielsweise nicht substantiell vom entsprechenden Grosshandelspreis; der komplette Wegfall aller physischen Materialkosten, Vertriebskosten, Lagerkosten, Personalkosten usw. müsste sich im Preis eigentlich deutlicher widerspiegeln als dies momentan der Fall ist (Man erinnere sich an die grossflächige Lancierung der Audio-CD, welche in der Produktion zwar wesentlich günstiger war als die früheren Audio-Kasetten, jedoch entgegen den Beteuerungen der Industrie zumeist zu einem höheren Preis als die entsprechenden Kasetten verkauft wurde). Auch die Frage, inwiefern pauschale, sogenannte 'Uhrheberrechtsabgaben' auf digitale Leermedien und

Kopierapparaturen bei einer Etablierung von DRM Systemen noch gerechtfertigt sind, ist derzeit Gegenstand von heftigen Diskussionen.

Es stellen sich zudem Fragen bezüglich Nachhaltigkeit und langfristiger Archivierug der (käuflich erworbenen) Inhalte. Was kann der Kunde beispielsweise unternehmen, wenn der bisherige Anbieter seiner zahlreich erworbenen DRM-Inhalte samt seinen proprietären Verfahren vom Markt verschwindet und eines Tages auch das Abspielgerät des Kunden ausfällt? Im schlechtesten Fall liegt schlussendlich nur noch eine verschlüsselte Datei vor, welche in dieser Form nutzlos ist; Der Aufwand, um die Verschlüsselung zu knacken, lohnt sich je nach eingesetztem Algorithmus wohl in den wenigsten Fällen. Überspitzt formuliert könnte man behaupten, dass der Kunde keine Gewähr hat, seine erworbenen Inhalte in Zukunft unabhängig vom entsprechenden Anbieter nutzen zu können.

Zu guter letzt setzen DRM-Agenten auch neue Anforderungen an die Hardware, was wiederum neue Beschaffungsinvesitionen für den Kunden bedeutet. Es sei hier jedoch angemerkt, dass gerade im Hinblick auf die Mobilfunkbranche technische Innovationen teilweise auch ohne die erhoffte entsprechende Kundennachfrage durch die Hersteller vorangetrieben wurden (Beispiel MMS). Wenn also der Kunde gar keine Wahl mehr hat, da sämtliche Endgeräte bereits mit DRM-Technologie ausgeliefert werden, dürfte sich diese Kritik relativieren. Die Fragestellung, ob sich Konzern- und branchenübergreifende Standards etablieren werden, oder ob schlussendlich doch der Markt mit proprietären Systemen fragmentiert wird, kann ohne weitere empirische Beobachtung der Marktentwicklung nicht beantwortet werden.

## 6.6 Zusammenfassung und Fazit

Zumindest gemäss Aussagen der Industrie wird in Zukunft kein Geschäft mit digitalen Inhalten ohne DRM-Mechanismen mehr möglich sein. Dies wird mit der Notwendigkeit von neuen Geschäftsmodellen (etwa Superdistribution oder Subskriptionsdienste) begründet, die sich nur mit DRM-Technologie umsetzen liessen. Besonders für mobile Geräte sei ein effektiver Schutzmechanismus für digitale Inhalte notwendig.

Diese DRM-Technologie befindet sich aber auch im Schnittfeld mit den Gesetzgebungen, insbesondere dem Urheberrecht und dem Datenschutzgesetz, welche nicht im Widerspruch zu DRM-Angeboten stehen dürfen. Eine internationale Betrachtung zeigt jedoch, dass die gesetzlichen Vorschriften zunehmends in Sinne der Industrie gelockert bzw. verschärft werden, etwa mit dem 'Digial Millenium Copyright Act' in den USA, welcher das Umgehen von technischen Kopierschutzmassnahmen verbietet.

Für DRM im mobilen Bereich sind vor allem die Spezifikationen der Open Mobile Alliance (ein Zusammenschluss von Firmen aus Mobilfunk und Informatik) für OMA DRM v1 und v2 bedeutsam, insbesondere der eingehend besprochene Anforderungskatalog. Derzeit sind jedoch noch kaum Implementierungen für OMA-DRM auf dem Markt, dafür verschiedene proprietäre Systeme wie etwa Microsoft Windows Media DRM, Shell Helix DRM oder Apple iTunes, welche sich mehr oder weniger an den Vorgaben der OMA-Spezifikationen

orientieren. All diese Produkte haben jedoch gemeinsam, dass sie einerseits nicht kompatibel zueinander sind, andererseits die Kunden mit zusätzlichen Massnahmen an sich und zusätzliche Komplementärprodukte (etwa Apple mit dem iPod) zu binden versuchen. Bereits sehen Marktbeobachter darin ein Zersplittern des Marktes in einzelne proprietäre Insellösungen zu Lasten eines einheitlichen Standards, was den Markterfolg von DRM insgesamt bedrohen würde.

Neben den bereits aufgeworfenen Fragen des Datenschutzes und der mangelhaften Standardisierung stellen sich auch grundlegende Fragen bezüglich der Kundenakzeptanz. Viele Kunden sehen sich durch DRM eingeschränkt und in ihren Rechten beschnitten. Der Erfolg der derzeit wohl verbreitetsten DRM-Plattform, iTunes von Apple, scheint der Industrie jedoch zumindest teilweise Recht zu geben. Ob sich aber jemals ein einheitlicher Standard durchsetzen wird und was das schlussendlich wirklich für den Konsumenten bedeutet, kann hier nicht abschliessend beantwortet werden. Der Markt wird es zeigen.

# Glossar

**Digital** Kodierung von Informationen nach diskreten Werten; steht im Gegensatz zur analogen Datenspeicherung. Digitale Informationen haben unter anderem die Eigenschaft, verlustfrei vervielfältigbar zu sein (z.B. CDs, Dateien, DVDs), während analoge Inhalte normalerweise bei jeder Vervielfältigung Qualitätseinbussen erleiden (Musikkasetten, VHS,..).

**DRM** Abkürzung für 'Digital Rights Management', Sammelbegriff für unterschiedlichste technische Verfahren, welche dem Schutz von immateriellen Inhalten durch seine Urheber bzw. Herausgeber dienen.

**Endgerät** Der Begriff meint das Gerät, auf welchem schlussendlich die Angebote eines Anbieters genutzt werden können. Dies kann beispielsweise ein Computer, ein Handy, ein portabler Musikspieler oder ein anderes Gerät sein.

**iPod** Tragbares Musikabspielgerät von Apple; Komplementärprodukt zum Apple iTunes Music Store (siehe iTunes).

**iTunes** Online-Shop für Musik, welcher von Apple mit einem DRM-System zum Schutz der Inhalte ausgestattet wurde.

**Kopierschutz** Verfahren, um die (unrechtmässige) Vervielfältigung analoger oder digitaler Medien zu verhindern.

**Kryptografie** Übersetzen von Daten in eine für den Menschen unlesbare Form mithilfe von Schlüsselkonzepten; der Umkehrvorgang ist i.d.R. nur mit dem entsprechenden Schlüssel möglich. Dient der Geheimhaltung von Daten bzw. der Übertragung von vertraulichen Inhalten über öffentliche Kanäle. Verschlüsselungstechnologien sind fester Bestandteil in den meisten DRM-Systemen.

**Mobile DRM** DRM mit Fokus auf mobile Endgeräte (Kleincomputer, Handies, tragbare Musikplayer usw.) und Portabilität von geschützten Inhalten.

**MP3** Beliebter und bekanntester Standard zur Kompression von digitaler Musik. Digitale Musik kann mit MP3 auf einen Bruchteil Ihrer Grösse geschrumpft werden, und dies mit kaum wahrnehmbaren Qualitätsverlust. Somit lässt sich die Information leichter und schneller übertragen, beispielsweise per Internet.

**OMA** Abkürzung für 'Open Mobile Alliance', ein Industriekonsortium mit vielen prominenten Mitgliedern aus der Mobilfunk- und Informatikbranche

**Trusted Device** Bestandteil der meisten DRM Konzepte; es handelt sich dabei um spezielle Hard- und Softwarebestandteile, welche vom DRM-Betreiber als sicher und zuverlässig im Hinblick auf die Durchsetzung seiner Rechte auf einem fremden Endgerät betrachtet werden.

**Verschlüsselung** siehe Kryptografie

# Literaturverzeichnis


[1] Bundesgesetz über das Urheberrecht und verwandte Schutzrechte (Urheberrechtsgesetz, URG) vom 9. Oktober 1992 [online], http://www.admin.ch/ch/d/sr/231_1/

[2] Bundesgesetz über den Datenschutz (DSG)) vom 19. Juni 1992 [online], http://www.admin.ch/ch/d/sr/235_1/index.html

[3] Open Mobile Alliance, 'OMA DRM Architecture, Version 2.0' [online], http://www.openmobilealliance.org/release_program/docs/CopyrightClick.asp?pck=DRM&file=V2_0-20041207-C/OMA-DRM-ARCH-V2_0-20040715-C.pdf, accessed 13.5.2005

[4] Real,'Helix DRM technical Description' [online], http://www.realnetworks.com/products/drm/description.html, accessed 18.5.2005

[5] Entwurf Bundesgesetz über das Urheberrecht und verwandte Schutzrechte (Urheberrechtsgesetz, URG)

[6] Erläuternder Bericht zur Änderung des Bundesgesetztes über das Urheberrecht und verwandte Schutzrechte (Entwurf 2004)

[7] DRM, DRM-Patente und Mobile DRM, Berlecon Research, Januar 2005

[8] Rosenblatt, Bill, 'GSM Association Rejects New DRM Patent License Terms' [online], http://www.drmwatch.com/standards/article.php/3502901, accessed 5.5.2005

[9] NZZ Online, 'Apple lanciert iTunes Schweiz'[online], http://www.nzz.ch/2005/05/10/vm/newzzE8K1I9YY-12.html, accessed 10.5.2005

[10] Wikipedia, 'FairPlay' [online], http://en.wikipedia.org/wiki/FairPlay, accessed 10.5.2005

[11] aaronsw.com, 'Behind the ITunes Music Store: A technical description of ITMS and FairPlay' [online], http://www.aaronsw.com/2002/itms/, accessed 30.4.2005

[12] Microsoft Corporation, 'A General Overview of Windows Media DRM 10 Device Technologies' [online], http://download.microsoft.com/download/3/a/f/3afb9301-5ade-4247-98ba-7a06efb75168/Introducing_Janus_and_Cardea.doc, accessed 12.6.2005

[13] Stallman, Richard, 'The right to read', Communications of the ACM (Volume 40, Number 2) [online], http://www.gnu.org/philosophy/right-to-read.html

[14] Open Mobile Alliance, 'DRM Requirements Document Version 2.0' [online], `http://www.openmobilealliance.org/release_program/docs/CopyrightClick.asp?pck=DRM&file=V2_0-20050614-C/OMA-RD-DRM-V2_0-20040716-C.pdf`, accessed 13.5.2005

[15] Brad Cox, 'Superdistribution', Wired Magazine 2.09, 1994 [online], `http://www.wired.com/wired/archive/2.09/superdis.html`, accessed 20.5.2005

# Kapitel 7

# Mobile Spiele - Ein Überblick über verschiedene Technologien

*Emanuel Giger, Urs Huber*

*Mobile Spiele erfreuen sich immer grösserer Beliebtheit. Unterstützt wird diese Tendenz durch den technischen Fortschritt auf mobilen Geräten und den rasant wachsenden Markt in diesem Bereich. Die vorliegende Arbeit widmet sich dem Thema "Mobile Spiele": Neben einer Einführung in die Thematik wird die wirtschaftliche Seite dieser Branche aufgezeigt, indem der Markt und die Marktteilnehmer genauer beleuchtet werden. Um die aktuellen Technologien auf dem Markt auf ihre Eignung für mobile Spiele besser beurteilen zu können, wird ein auf Möglichkeiten und Problemen basierter Kriterienkatalog entwickelt, der es in einem späteren Schritt erlaubt, die heute zugänglichen Technologien und ihr Zukunftspotential zu beurteilen.*

## Inhaltsverzeichnis

**Glossar**

| Begriff | Bedeutung |
|---|---|
| **2G** | Zweite Generation von kabellosen Services, wie CDMA, TDMA und GSM |
| **3G** | Dritte Generation von kabellosen Services. Dazu gehört unter anderem UMTS |
| **Actionspiel** | Videospiel mit starker Betonung des Echtzeit-Aspektes. Actionspiele sind meist so gestaltet, dass die Zeitperioden, die dem Spieler für Reaktion und Planung zur Verfügung stehen, im Sekundenbereich liegen |
| **API** | Application Programming Interface |
| **Beat'em Up** | Die Spieler müssen versuchen, den Gegner durch den Einsatz von körperlicher Gewalt in Form einer Kampfsportart zu besiegen. |
| **BREW** | Binary Runtime Environment for Wireless |
| **CDC** | Connected Device Configuration |
| **CDMA** | Code Division Multiple Access |
| **CLDC** | Connected Limited Device Configuration |
| **DRM** | Digital Rights Management |
| **GAPI** | Game API |
| **GSM** | Global System for Mobile Communication |
| **I/O** | Bezeichnet Input/Output eines Computersystems |
| **J2ME** | Java2 Micro Edition |
| **JNI** | Java Native Interface |
| **JRE** | Java Runtime Environment |
| **JVM** | Java Virtual Machine |
| **Jump'n Run Spiel** | Videospiel, bei denen sich die Spielfigur laufend und hüpfend zu einem bestimmten Ziel fortbewegt |
| **MIDP** | Mobile Information Device Profile |
| **MMOG** | Massive Multiplayer Online Games |
| **MMS** | Multimedia Messaging Service |
| **NTT** | Nippon Telegraph and Telephone Corporation ist in Japan der Marktführer unter den Telekommunikationsunternehmen |
| **OS** | Operating System bzw. Betriebssystem |
| **OTA** | Over-The-Air |
| **PDA** | Personal Digital Assistent |
| **Rollenspiel** | In einem Rollenspiel übernehmen die Spieler die Rollen von fiktiven Charakteren, die in einer meist fantastischen Welt Abenteuer erleben |
| **SDK** | Software Development Kit |
| **SIM** | Subscriber Identity Module. Chipkarte für das Mobiltelefon |
| **SMS** | Short Message Service |
| **SVGA** | Super Video Graphics Array |
| **TDMA** | Time Division Multiple Access |
| **TFT** | Thin Film Transistor |
| **UMTS** | Universal Mobile Telecommunication System |

**Fortsetzung Glossar**

| **Begriff** | **Bedeutung** |
|---|---|
| **WAP** | Wireless Application Protocol. Ein Protokoll für die mobile Telekommunikation |
| **WLAN** | Wireless Local Area Network |

## 7.1 Einführung

Der Markt für mobile Spiele in Europa steckt noch in der Kinderschuhen, verspricht aber für die Zukunft ein hohes Potential mit enormen Wachstumschancen. Dies wird deutlich, wenn der europäische Markt mit dem Asien-Pazifik Markt verglichen wird, der bereits jetzt alleine mit dem Verkauf von Handyspielen einen Umsatz von 240 Millionen erwirtschaftet und bis im Jahr 2008 bis zu 1,3 Mrd. Euro erwirtschaften soll [2].

Bis diese Marktgrösse erreicht wird, müssen noch einige Probleme einerseits auf betriebswirtschaftlicher Seite (z. B. die Aufteilung der Gewinne entlang der Wertgenerierungskette oder das Bereitstellen von Spielen auf verschiedenen Portalen, etc.) andererseits auch auf technischer Seite gelöst werden. Dazu gehören insbesonders die Entwicklung eines technologischen Standards, welcher von den meisten mobilen Geräten unterstützt wird, um das optimale Verhältnis zwischen Aufwand und Ertrag für die Entwickler wie auch für die Publisher zu erzielen.

In dieser Arbeit wird eine top-down Strategie zur Aufarbeitung des Themas gewählt, indem im ersten Kapitel zuerst aktuelle und zukünftige Anwendungsszenarien aufgezeigt werden, bevor dann auf die Terminologie des Begriffs Mobile Spiele eingegangen wird. Danach wird aus einem wirtschaftlichen Standpunkt die Grösse und das Potential des mobilen Marktes in Europa aufgezeigt, und mit denen auf anderen Kontinenten verglichen. Ausserdem werden die Marktteilnehmer genauer beleuchtet und ihre Interessen bzw. Strategien erläutert. Um die zukünftige Entwicklung besser verstehen zu können ist es nötig, einen genaueren Blick auf die verfügbare Technologie zu werfen und die Anforderung an sie, die Probleme und Möglichkeiten dieser aufzuzeigen. Diese Ansatzpunkte führen dann zu verschiedenen Kriterien, welche in einen Kriterienkatalog einfliessen. Diese Kriterien sind für einen zukünftigen Erfolg der mobilen Spiele aus Sicht der Technik nötig. Das abschliessende Kapitel betrachtet die bis anhin verfügbaren Technologien und zeigt auf, welche Kriterien bereits erfüllt sind bzw. bei welchen noch Aufholbedarf besteht.

## 7.2 Einführung in die mobilen Spiele

In diesem Kapitel werden mögliche Anwendungsszenarien beschrieben und zukünftige Möglichkeiten der mobilen Spiele aufgezeigt. Danach erfolgt eine Definition der mobilen Spiele mittels einer Klassifikation.

### 7.2.1 Anwendungsszenarien und Möglichkeiten

Die Anwendungsszenarien und Möglichkeiten werden nach den Spieltypen - Einzelspieler- und Multiplayerspiele - unterschieden.

**Einzelspieler Spiele**

Für aktuelle mobile Geräte (wie Mobiltelefone und PDA) sind momentan verschiedene Klassen von Spielen verfügbar [35]. Eine der ersten Klassen für mobile Geräte waren die eingebetteten Spiele - Spiele, welche auf dem Chip der mobilen Geräte installiert sind. Ein prominentes Beispiel ist Snake, welches auf den ersten Nokia Mobiltelefonen verfügbar war.

Eine andere Möglichkeit stellen die SMS-Games dar, wobei der Benutzer eine Textnachricht an den Anbieter schickt, welcher dann die empfangene Nachricht auswertet und eine Antwortnachricht zurückschickt. Durch die Beschränkung auf textuelle Nachrichten ist diese Technolgie für Spiele eher ungeeignet. Dagegen kann mittels der MMS-Technologie diese Art von Spiele attraktiver, aber von Prinzip her nicht interessanter gemacht werden.

Eine weitere Technologie, welche dem Spieler ermöglicht auf mobilen Geräten zu spielen, ist WAP (Wireless Application Protocol). Anwender können die sogenannten Browsing Games spielen, indem sie mittels WAP auf die URL der Anbieter gehen, Seiten betrachten, sich durch Menus klicken oder Texteingaben machen, diese Daten dem Server übermitteln, um dann entsprechend weitere Seiten anzuschauen. Durch die statische Technologien ist die Interaktivität dieser Spiele beschränkt.

Gegenwärtig können Spiele auf das mobile Gerät, falls unterstützt, heruntergeladen werden, um sie dann lokal auf dem eigenen mobilen Gerät spielen zu können. Dabei kann der Anwender aus einer breiten Palette von Genres auswählen, zu diesen gehören unter anderem rundenbasierte Strategie-, Jump'n Run-, Action-, Sport-, wie Geschicklichkeitsspiele. Jedoch befindet sich die Qualität der Spiele aufgrund der begrenzten Hardware im Entwicklungsstand der PC-Industrie im Jahre 1995, als 3-D Beschleunigerhardware noch ein Fremdwort war, und die Rechner- und Speicherkapazität sich auf wenig anspruchsvolle Anwendungen beschränkte. Betrachtet man nun die rasante Entwicklung der Hardware der Mobiltelefone und der PDAs und die im Gleichschritt verlaufende Verbesserung der angebotenen Spiele, so kann man damit rechnen, dass in ein paar Jahren mobile Spiele auf mobilen Geräten eine annähernde Qualität wie heutige PC- und Konsolenspiele haben könnten.

**Multiplayer Spiele**

Eine weiteres Anwendungsszenarium stellen die Multiplayerspiele dar. Das Potential von Multiplayerspielen kann am Beispiel von PC-Onlinespielen wie Everquest2 und World of Warcraft abgeschätzt werden. So wurden z.B. schon am ersten Verkaufstag des Online-Rollenspiels World of Warcraft alleine in Deutschland 280'000 Exemplare verkauft [64]. Im Gegensatz zum PC besteht für die mobilen Geräte schon heute die Möglichkeit interaktiv mit anderen Personen bzw. Geräten zu kommunizieren, da sie schon von Natur aus verbunden sind, sei es nun über das 2G oder 3G-Netz, WLAN oder Bluetooth. Bis heute ist es z.B. aufgrund der beschränkten Bandbreite (Latenz) und der Leistungsfähigkeit der mobilen Geräte nur beschränkt möglich netzwerkbasiert zu spielen. So können z.B. nur Punktestände via SMS/MMS oder WAP ausgetauscht werden. Eine andere bereits

heute verfügbare Möglichkeit ist der Austausch zwischen Spielern über Bluetooth, welches ad-hoc Netze mit einer maximalen Anzahl von 8 Spieler auf einem begrenzten Raum ermöglicht [38].

Um aber Massive Multiplayer Online Games (MMOG) zu entwickeln, in denen tausende von Personen zusammengeführt werden, um sich zu unterhalten, gemeinsam Aufgaben auszuführen oder gegeneinander anzutreten, muss die Entwicklung der Technologie, insbesondere der Übertragunsmöglichkeiten von Daten, weiter vorangetrieben werden.

### 7.2.2 Einordnung der mobilen Spiele

Das mobile Spiele Marktsegment kombiniert den Videospielsmarkt aus der Unterhaltungsindustrie mit den mobilen Datenapplikationen aus der Telekommunikationsindustrie. Dabei gehören zum Videospielmarkt als wichtigster Vertreter das PC- und Konsolenspiele-Segment. Dazu gehören auch bestimmte tragbare Spielkonsolen wie der Nintendo Gameboy [12] und Sonys PlayStationPortable (PSP) [51]. Auf diese wird in dieser Arbeit nicht eingegangen, da ihre Betriebssysteme geschlossen sind und nur Spiele darauf gespielt werden können, die speziell für diese Plattform entwickelt wurden. Das mobile Datenapplikationen-Segment enthält die meisten nicht-sprachlichen Dienstleistungen, wie z.B. SMS, MMS, Internet und Intranetzugang wie auch Informationsbereitstellungsservice, welche von einem Telekommunikationsanbieter angeboten werden [15].

Ein Übersicht über die Verbindung zwischen diesen zwei Märkten und den drei Hauptelemente des mobilen Spielens ist in Abbildung 7.1 ersichtlich.

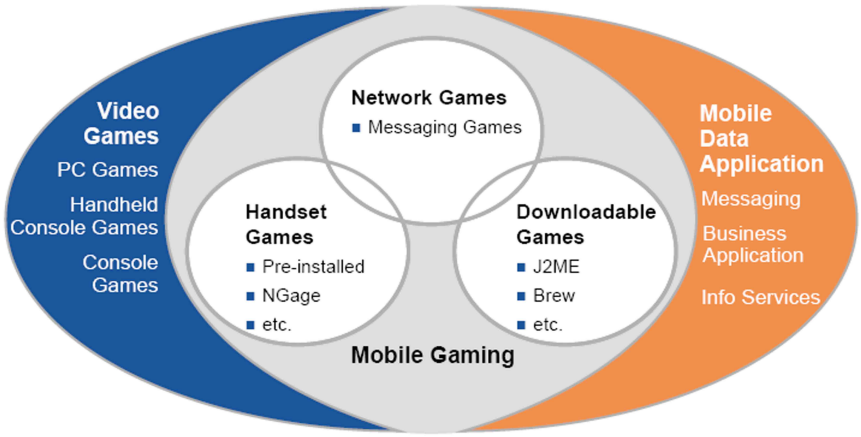


Abbildung 7.1: Mobile Spiele als Verbindung zwischen Video Spielen und mobilen Datenanwendungen [15]

Alle mobilen Spiele gehören zum Videospielmarkt, sind aber auch durch ihren Einsatzort (auf mobilen Geräten), ihrer Distribution (herunterladbar oder vorinstalliert) und ihren Interaktionsmöglichkeiten mit anderen Spieler (Netzwerk) auch Teil des Telekommunikationsmarkts und können somit von den Kommunikationsmöglichkeiten profitieren.

**“Mobile Geräte“-Spiele**

Zu dieser Art von Spielen gehören einerseits die eingebetten Spiele, welche auf dem mobilen Gerät vorinstalliert sind und andererseits Spiele, welche separat auf einem Speichermedium gekauft und dann auf dem mobilen Gerät gespielt werden können, dazu gehört z.B. Nokia N-Gage [15].

**Netzwerk-Spiele**

Netzwerkspiele benötigen Netzwerkstandards und -protokolle, um Spiele auszuführen. Dieses Segment kann weiter in zwei Unterkategorien unterteilt werden: Auf der einen Seite sind dies Messaging Games, wie z.B. SMS, MMS und WAP Games und auf der anderen Seite die interaktiven Spiele. Beispiele für interaktive Spiele sind: Die Möglichkeit den Punktestand auf ein Portal heraufzuladen, Multiplayerspiele bis hin zu ihrer extrem Ausprägung der Massive Multiplayer Online Games und anderer Services [15].

**Herunterladbare Spiele**

Herunterladbare Spiele werden über das mobile Netzwerk vertrieben und auf dem mobilen Gerät installiert und gespielt. Die unterstützten Geräte sind dabei Handys, PocketPCs oder auch Palms. Während der letzen Jahre hat sich vorallem die Java2 Micro Edition (J2ME) Plattform auf dem europäischen Markt, die Binary Runtime Environment for Wireless (BREW) Plattform auf dem amerikanischen Markt und die GVM Plattform auf dem asiatischen Markt durchgesetzt. In dieser Arbeit gehen wir insbesondere auf die herunterladbaren Spiele ein, da sie einerseits einen sehr grossen Einfluss auf das mobile Spielen haben, und andererseits die Position solcher Spiele im Gegensatz zu den anderen Arten mobiler Spiele in Zukunft weiter gestärkt wird [15].

### 7.2.3 Evolution der mobilen Spiele

Um die Entwicklung bzw. die Evolution der mobilen Spiele zu verdeutlichen, zeigt Abbildung 7.2 eine Übersicht über eine mögliche Entwicklung der mobilen Spiele.

Momentan befindet sich die Entwicklung etwa in der Mitte der Abbildung 7.2, d.h. akutelle werden Blockbuster aus dem PC- und Konsolenbereich, wie z.B. Splinter Cell und Anno 1503, auf mobile Geräte adaptiert. Es wird aber nicht mehr lange dauern, bis weitere qualitativ gute 3D-Spiele auf den Markt kommen. Die Entwicklung von Mobile Massive Multiplayer Rollenspielen, welche aufgrund der Interaktivität und der Breite des Spielinhalts komplex sind, wird erst in den nächsten Jahren möglich sein.

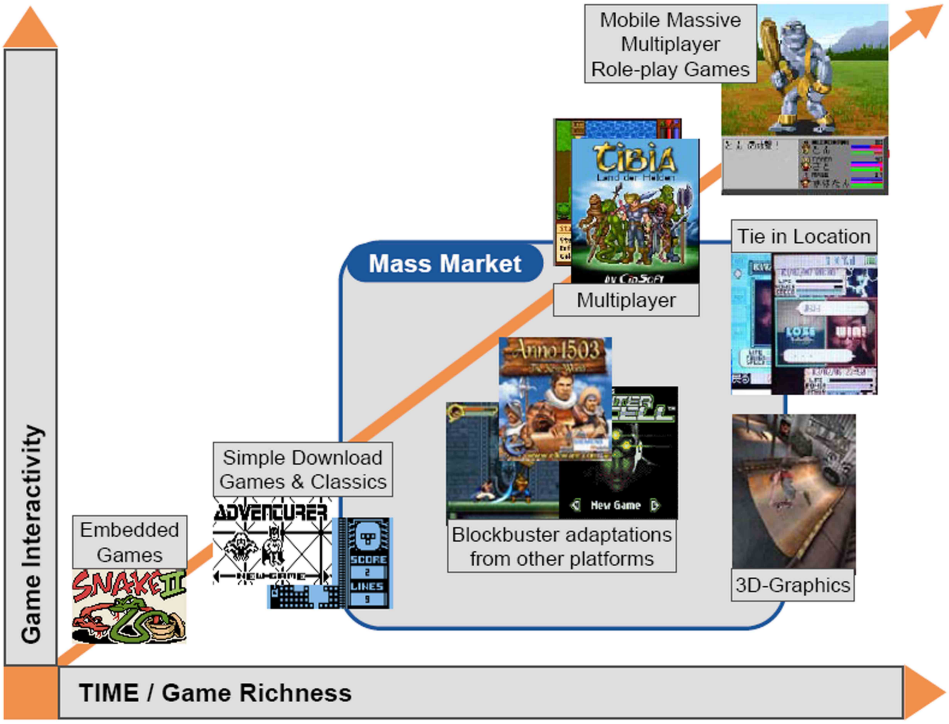


Abbildung 7.2: Die Evolution der mobilen Spiele, Quelle [15]

## 7.3 Der mobile Spiele Markt

In diesem Kapitel wird auf die Möglichkeiten des Marktes mobiler Spiele eingegangen und seine Entwicklung in regional unterschiedlichen Märkten verglichen. Ausserdem werden verschiedene Hersteller mobiler Geräte und Spiele vorgestellt.

### 7.3.1 Markt in Europa: Heute und Morgen

Wie in der Einführung beschrieben können mobile Spiele und speziell das Herunterladen von Spielen auf mobile Geräte einen multimilliarden Markt eröffnen. Eine Übersicht über die bisherigen und zukünftigen Einnahmen aus dem Mobile Spiele Markt von Frost and Sullivan aus der Studie European Mobile Gaming ist auf Abbildung 7.3 ersichtlich.

Demnach werden die Gesamteinnahmen von 3 Mrd. Dollar in 2004 sich bis ins Jahre 2006 auf 6 Mrd. Dollar verdoppeln. Insbesondere sind 75% der Einnahmen im Jahre 2006 auf herunterladbare Spiele zurückzuführen.

Wie in der Abbildung 7.3 zu sehen ist, erwirtschaft der Sektor der Netzwerkspiele, d.h. das Messaging, WAP und interaktive Spiele, aktuell nur einen Bruchteil der gesamten Einnahmen. In Zukunft aber wird vorallem auch der Sektor der interaktiven Spiele an mehr Einfluss gewinnen. Wenn technisch mehr Multiplayerspiele möglich sind, dann können z.B. aufgrund der Notwendigkeit einer zentralen Infrastruktur (Server), indem sich die Spieler via den mobilen Geräten treffen können, dem Spieler periodische Zahlungen abverlangt

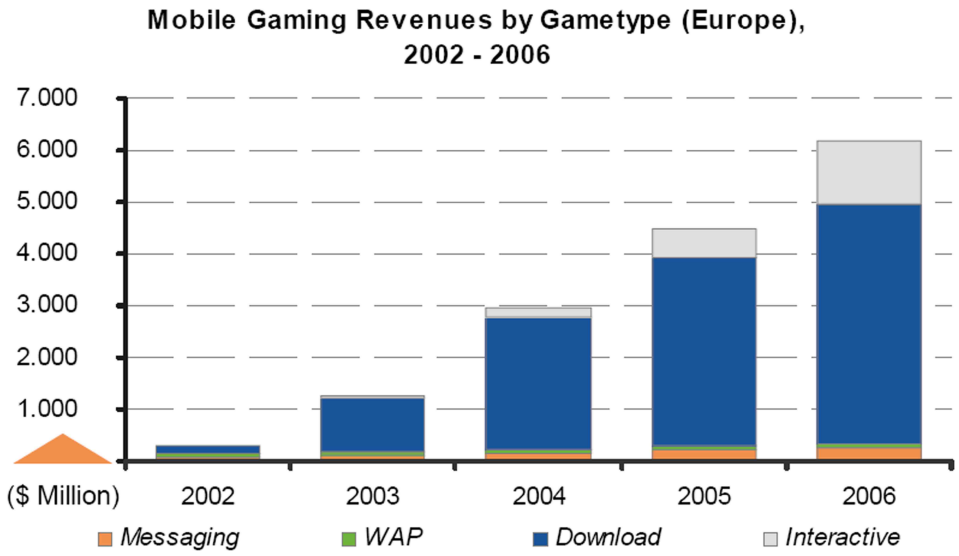


Abbildung 7.3: Einnahmen aus mobile Spiele Markt in Europa, Quelle [15]

werden, vorausgesetzt ein Geschäftsmodell bzw. die Zahlungsmodelle lassen dies zu. So würden Netzwerkspiele im Gegensatz zu den Einzelbenutzerspielen fortlaufend Einnahmen generieren [36].

### 7.3.2 Markt in Japan: Eine Erfolgsgeschichte

Im Gegensatz zum europäischen Markt, der immer noch relativ klein ist, ist der asiatische Markt bzw. der japanische und südkoreanische Markt erfolgreich. Nach Analysen entfallen ca. 80-90% der aktuellen globalen Einnahmen der mobilen Spiele Branche auf diese beide Märkte [34]. Die Mobile-Content-Industrie in Japan setzt heute bereits rund drei Milliarden Dollar im Jahr um, wobei die Einnahmen der Netzbetreiber durch Datenverkehr noch nicht berücksichtigt sind. Insgesamt ist Japan der weltweit umsatzstärkste und technologisch am weitesten entwickelte Mobilfunktmarkt [14].

Europa steht im Vergleich dazu erst an der Schwelle zu einer solchen Entwicklung. Zwar bietet Mobile Media auch für europäische Mobilfunkanbieter ein hohes Ertrags- und Wachstumspotenzial, doch agieren die Anbieter noch nicht gemeinsam mit den traditionellen Medienunternehmen. Warum ist der japanische Markt so erfolgreich?

- Viele der erfolgreichen Spiele in Japan sind **multiplayer Spiele**. Ein Beispiel für ein solches Spiel ist das Samurai Romanesque Spiel, welches über die Spielentwicklerseite i-mode (i-mode ist ein dem Internet ähnlicher mobiler Dienst für Handys [18]) für 3 Euro pro Monat vertrieben wird. Das Spiel ist ein komplettes multiplayer Spiel basierend auf einer Java Applikation und ermöglicht dem Spieler neben dem Bekämpfen anderer Spieler auch das Dating mit diesen [15].

- Die **Integration von Nachrichten und Kommunikationsmöglichkeiten der mobilen Geräte** in Spielen ist ein weiterer Grund. So hat z.B. ein Spiel eine Mailingfunktion integriert, die es ermöglicht anderen Spielern eine Nachricht zu schicken. Diese Möglichkeit hat auch einen positiven Effekt auf den Gebrauch der Nachrichtenservices [15].

- Eine der Schlüsselgründe für den Erfolg von i-mode in Japan ist, dass NTT DoCoMo (Mobilfunk-Ableger der NTT in Japan und vergleichbar mit T-Mobile in Deutschland [39]) **klare Spezifikationen für ihre mobilen Geräte** definierte. Dies ermöglicht es Applikationsentwicklern für einen breiteren Markt zu entwickeln, ohne zusätzlichen Programmier- und Testaufwand in Kauf zu nehmen [15].

- NTT DoCoMo unterstützt ausserdem die **unabhängigen Inhalteanbieter mit einer kostenlosen Entwicklungssoftware und der Bereitstellung eines Billingsystems**. So konnten innerhalb kurzer Zeit die kritische Masse von 2,5 Mio. i-mode Nutzern gewonnen werden, auf die rasch weitere Anbieter und Nutzer folgten [14].

### 7.3.3 Zielgruppe

Grundsätzlich können Spieler in Europa in zwei Segmente aufgeteilt werden: Zum einen die Spielfanatiker bzw. Hardcore Gamers und die Gelegenheitsspieler bzw. Casual Gamers. Spielfanatiker sind Leute, die mehr als 10 Stunden pro Woche spielen. In Europa gibt es ca. 1,2 Mio. Spielefanatiker, typischerweise männliche Teenager. Dieses Segment kann wieder in zwei Subsegmente unterteilt werden: Die Spieler, welche vom traditionellen Spielen kommen (wie PC und Konsolen Spieler) und Spieler, welche vom alltäglichen mobilen Spielen her kommen [48]. Eine Übersicht über die Zielgruppensegmente zeigt die Tabelle 7.1.

Das zeigt, dass Spielfanatiker nur dann auf mobile Spiele umsteigen, wenn diese Spiele ein ähnliches Erlebnis wie PC- bzw. Konsolenspiele liefert, d.h. verschiedene Zielgruppen haben verschiedene Erwartungen vom mobilen-Spiele Erlebnis.

Für Spieleentwickler und -publisher bedeutet das, dass vorerst vorallem die Gelegenheitsspieler bedient werden - die sogenannte M-Generation, d.h. Kinder und junge Erwachsene im Alter zwischen 15 und 29. Marktanalysten zeigten, dass die generelle Nachfrage nach mobiler Unterhaltung (wie mobile Spiele, Musik, SMS,...) mit zunehmendem Alter über diesem Intervall abnimmt. Insbesondere gilt dies für die mobilen Spiele, welche die höchste Nachfrage bei der M-Generation haben und sehr schnell an Attraktivität im älteren Teil des Segment verlieren [15]. Für die Spieleentwickler bedeutet dies, dass sie Spiele entwickeln müssen, die einerseits kurze Spielesessionen erlauben und andererseits auch die Speicherung dieser ermöglichen.

Erst wenn die technischen Einschränkungen weiter abnehmen, und somit die Entwicklung von grafisch und inhaltlich komplexeren Spiele möglich wird, verschiebt sich die Zielgruppenverteilung zwischen Gelegenheitsspielern und Spielefanatikern weiter zu Gunsten der Spielefanatiker.

Tabelle 7.1: Mobile Spieler Segmentation, in Anlehnung an [48]

| **Eigenschaften** | **Mobile Gelegenheitsspieler** | **Mobile Spielfanatiker** | |
|---|---|---|---|
| | | Vom alltäglichen mobilen Spielen | Vom traditionellen hardcore Spielen |
| **Demographie** | Jung, beide Geschlechter & jedes Alter und beide Geschlechter | Jung und beide Geschlechter | Jung und männlich |
| **Warum mobile Spiele?** | Zeitvertreib | Hobby | Hobby |
| **Typ des Spiels** | Einfach | Einfach & Komplex | Komplex |
| **Erfolgsfaktor** | Gameplay | Gameplay & Grafik | Grafik |
| **Anteil der Bevölkerung** | Mehrheit | Minderheit | Minderheit |
| **Ab wann wird Zielgruppe bedient?** | Jetzt | Abgestuft | Falls genug gute mobile Geräte verfügbar |

### 7.3.4 Schlüsselanbieter

Um die Schlüsselanbieter überhaupt identifzieren zu können, muss zuerst die mobile Technologie in ihre Subfelder unterteilt werden. Dazu liefert die Abbildung 7.4 eine gute Übersicht:

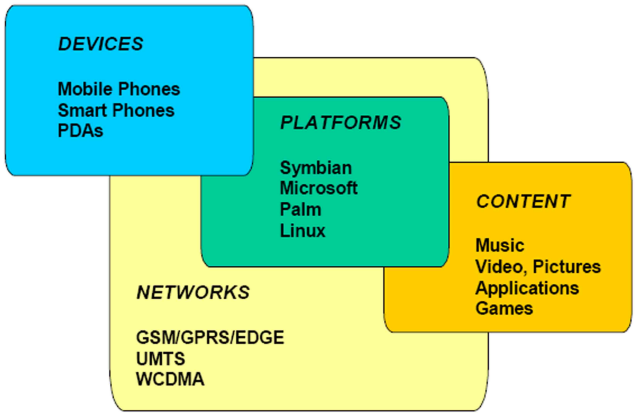


Abbildung 7.4: Technologie für das mobile Spielen [30]

#### Gerätehersteller

Wie in Abbildung 7.4 gezeigt, zählen zu den mobilen Geräte die Mobiltelefone, die Smartphones und die Personal Digital Assistens (PDAs). Hersteller von mobilen Geräten gibt

es verschiedene, wobei nachfolgend nur auf die wichtigsten mit ihrer strategischen Ausrichtung eingegangen wird:

- **Nokia**: Das finnische Unternehmen Nokia ist wohl in Europa der bekannteste Hersteller von Mobiltelefonen, Smartphones und Communicators. Mit einem Marktanteil von 34.4 % ist Nokia die klare Nummer Eins im europäischen Markt für Mobiltelefone [16]. Daneben produziert Nokia Digitalreceiver für digitales Fernsehen und PCMCIA-Mobilfunkkarten, welche einem Laptop oder PDA den Zugang zu Netzen ermöglichen. Nokia fokussiert nicht auf ein bestimmtes Kundensegment, sondern möchte alle möglichen Kundengruppen bedienen:
  - **Jugendliche**: Mittels dem mobilen Spielgeräts N-Gage
  - **Privatnutzer**: Mit verschiedensten mobilen Geräten, welche Zusatzfeatures wie digitale Kamera, Radio, Videorecorder und Java-Unterstützung beinhalten - je nach Vorlieben des Kunden.
  - **Geschäftsnutzer**: Einerseits durch Mobiltelefone, die Triband unterstützen und Videotelefonie ermöglichen, und andererseits Smartphones und Communicator, mit kompletter Tastatur, grösseren Displays für E-Mail Abfragen und HTML-Browsing und Office-Programmen (wie Organizer, etc.).

  Dies wird auch an der Anzahl der aktuell verfügbaren Geräten von 52 widerspiegelt. Die sehr breite Produktpalette rundet Nokia mit einem Online-Angebot von Sounds, Grafiken, Games und Applikationen ab. Durch die Entwicklung von anderen Geräten, wie Digitalreceiver und PCMCIA-Karten, möchte Nokia auch in andere Marktsegmente vorstossen. [37]

- **Siemens**: Der zweitgrösste Hersteller von Mobiltelefonen mit einem Marktanteil von 15.2% ist Siemens [16]. Siemens bietet eine breite Palette von Produkten, Lösungen und Services an - Von Automatisierungssysteme bis zur Verkehrstelematik. Im Bereich der mobilen Geräte sind es Mobiltelefone, Smartphones wie auch PDAs bzw. Pocket-PCs. Dabei liegt der Fokus vorallem auf Geschäftsleute, aber auch immer mehr auf Privatpersonen. Junge Mobiltelefonnutzer finden bei Siemens durch das eingeschränkte Angebot in diesem Sektor wenig Beachtung. Ausserdem betreibt Siemens nur ein sehr kleines Online-Portal für das Herunterladen von zusätzlichem Inhalt. Nur Bilder und Klingeltöne werden angeboten [50][1].

- **Sony Ericsson**: Sony Ericsson ist der drittgrösste Anbieter von Mobiltelefonen in Europa. Sony Ericsson bietet Mobiltelefone und vereinzelt Smartphones an. Neben der teilweisen Integration eines digitalen Musikplayers überzeugen die Geräte vorallem durch ein sehr grosses Display und sind daher unter den Mobiltelefonen am besten als Fotoapparat geeignet. Diese multimediale Ausrichtung lässt darauf schliessen, dass Sony Ericsson vorallem Privatnutzer jüngeren Alters ansprechen will, die das Mobiltelefon nicht nur als Telekommunikationsmittel verstehen. Die geschäftlichen Nutzer werden durch die wenig angebotenen Smartphones nur unvollständig bedient. Im Gegensatz zu Siemens bietet Sony Ericsson ein Online-Portal mit Bildern, Klingeltönen, Games, Applikationen und Tools an [52].

[1] Mitte 2005 übernahm die Firma BenQ das Mobiltelefongeschäft von Siemens [4].

- **Palm**: Palm ist momentan der Marktleader im europäischen PDA-Markt mit einem Marktanteil von 42% [11] vor Hewlett Packard, welche einen Marktanteil von 36% hat. Neben der Herstellung von PDAs stellt Palm seit Ende 2002 auch Smartphones her. Die angebotenen Geräte zeigen deutlich, dass der Fokus auf Geschäftsleute ausgerichtet ist, da keine Mobiltelefone hergestellt werden, sondern vorallem Computer im Kleinformat das Ziel sind. So unterstützen die Geräte Navigationslösungen, den einfachen E-Mail Abruf und eine breite Palette von mehr als 15'000 verschiedener Softwarelösungen (wie z.B. Programme zur Anzeige und Bearbeitung von Office-Dateien). Diese können im Online-Portal auch direkt gekauft und über den Download via eines PC auf den PDA geladen werden. Mit dem Einstieg in den Mobilfunkmarkt mittels den Smartphones will Palm den Spagat zwischen den hoch entwickelten PDAs und der Unterstützung der Mobiltelefonie schaffen, um so neue Kunden mit gehobenen Ansprüchen zu gewinnen [42]. Neben der Herstellung von PDAs stellt Palm auch das Betriebssystems für Palms her - PalmSource. Dieses wird vorallem in den USA verwendet [43].

### Plattformen

Zu den wichtigsten Herstellern von Betriebssystemen gehören - wie in Abbildung 7.4 ersichtlich - Symbian, Microsoft, Palm und Linux. Auf Palm und Linux wird nicht detaillierter eingegangen, da sie einen zu kleinen Marktanteil aufweisen. Neben den Betriebssystemen sind die eine Schicht höher liegenden Applikationsplattformen von Sun (Java) und Qualcomm (BREW) relevant. Eine eingehendere Analyse relevanter Betriebssysteme und Applikationsplattformen folgt in Kapitel 7.5.

### Spieleanbieter

Es existieren weit über 100 Entwickler für mobile Spiele. Da im europäischen Markt vorallem Java-Spiele verbreitet sind, wird nachfolgend auf drei der grösseren Entwicklerfirmen eingegangen, die mit der J2ME Technologie arbeiten.

- **Elkware**: Der grösste deutsche Hersteller Elkware wurde Ende 2004 von der Internet-Holding InfoSpace aufgekauft. Im dritten Quartal 2004 hat Elkware 1 Mio. Dollar erwirtschaftet [19]. Elkware bietet aktuell 80 verschiedene Spiele in verschiedenen Genres an, darunter sind auch Umsetzungen von PC- und Konsolenspielen, wie das Rollenspiel Sacred oder den Strategiespielen The Fall, Anno 1503 und Port Royal 2. Bei der Entwicklung der Spiele setzt Elkware vollständig auf Java (J2ME) [10]. Ziel ist, aktuelle Blockbuster aus der PC- und Konsolenspielewelt mittels Lizenzierung auf mobile Geräte zu portieren, um eine möglichst grosse Spielergemeinschaft für sich zu gewinnen.

- **Gameloft**: Die amerikanische Firma Gameloft, welche 1999 gegründet wurde, stellt Spiele für mobile Geräte auf Basis von Java, Brew und Symbian her. Die erwarteten Verkaufszahlen im Jahre 2007 sollten bei 1 Mia. verkaufter Spiele liegen. Gameloft

hat viele Partnerschaften mit Lizenzgebern wie Ubi Soft und ist berechtigt Markennamen von Spielen solcher grosser Hersteller weiterzuverwenden [13]. Dazu gehören z.B. die auf PC- und Konsolenbereich erfolgreichen Spiele wie Rayman, Tom Clancy's Splinter Cell, Ghost Recon, Prince of Persia und Tom Clancy's Rainbow Six.

- **Macrospace**: Macrospace bietet neben mobilen Spielen auch mobile Spielelösungen für Netzwerkoperatoren, Serviceprovider und Gerätehersteller an. Die Spiele werden auf verschiedenen Plattformen entwickelt, dazu gehören Java, BREW, Symbian, DoJa und Mophun und können somit eigentlich alle potentiellen Spieler erreichen. Das Unternehmen Macrospace ist ein Pionier im Bereich connected gaming. Macrospace offeriert online High Score, in-game Event Billing Lösungen, mobile Chat und Anwenderforen - ein erster Schritt bei der Entwicklung von Multiplayerspielen. Zu den bekanntesten Spielen von Macrospace gehört das rundenbasierte Strategiespiel Ancient Empires [29].

## 7.4 Kriterien für ein erfolgreiches mobiles Umfeld

Mobile Gaming ist eine relativ neue Art der elektronischen Unterhaltung. Das Phänomen "Mobile Gaming" ist in erster Linie auf die technische Entwicklung bei den mobilen Endgeräten und auf ein verändertes Kommunikationsverhalten der heutigen Gesellschaft zurückzuführen. So betrug die Anzahl aktiver SIM Karten in der Schweiz im Dezember 2001 5'341'000 [26]. In diesem Kapitel werden daher Probleme, Schwierigkeiten aber auch Vorteile und Möglichkeiten untersucht, die entscheidend sind, ob sich Mobile Gaming zu einem Massenmarkt entwickelt oder nicht. Ziel ist es, einen Kriterienkatalog zu erstellen, der die Eigenschaften von Mobile Gaming aufzeigt. Mit Hilfe dieses Katalogs werden im folgenden Kapitel die vorherrschenden Technologien auf dem Markt auf ihre Eignung für Mobile Spiele untersucht und beurteilt. Im folgenden Abschnitt wird zwischen technisch bedingten und nicht technisch bedingten Eigenschaften unterschieden.

### 7.4.1 Technisch bedingte Eigenschaften

Obwohl die technische Entwicklung bei mobilen Endgräten entscheidende Fortschritte machte, liegen die wichtigsten Probleme immer noch in den technischen Grenzen der Geräte, die heute im mobilen Umfeld benutzt werden.

#### Display, Farben und Sounds

Eine grosse Einschränkung stellen die kleinen Displays und ihre Qualität in Bezug auf Auflösung und Farbtiefe dar. Viele Geräte, die heute benutzt werden, besitzen immer noch einen schwarz-weiss Display. Auch die Wiedergabe von Tönen ist auf den meisten Geräten noch bescheiden. Diese Restriktionen stellen eine grosse Herausforderung für Spielentwickler dar, da Mobile Gaming eine audiovisuelle Tätigkeit ist, und die Präsentation ein wichtiger Faktor für die Qualität eines Spiels darstellt. Zumindest einige dieser

Einschränkungen werden durch die neue Generation von mobilen Endgräten entschärft. So besitzen zum Beispiel viele neue Mobilefunktelefone Farbdisplays, Fotokameras und sind in der Lage komplexe Sounds abzuspielen und zeigen die Tendenz zu Multimediageräten auf: Zum Beispiel das neue Nokia 8800 mit einer SVGA Kamera und einem internen Aktivmatrix-TFT-Farbdisplay mit 262‘144 Farben und einer Auflösung 208 x 208 Pixels [37]. Die limitierte Grösse des Displays liegt in der Natur der mobilen Geräte und wird trotz der technischen Entwicklung immer eine Einschränkung bleiben. Hier sind in erster Linie die Entwickler von mobilen Spielen gefordert. Sie müssen Spielprinzipien entwickeln, die auch auf kleinen Displays funktionieren.

### Inputs and Controls

Die meisten mobilen Endgeräte wurden nicht primär für das Mobile Gaming entwickelt und gestaltet, sondern in erster Linie als mobile Kommunikationsgeräte. Die Geräte besitzen daher nur ein einfaches Tastenfeld, welches ausserdem noch sehr klein ist. Spiele erfordern aber oft vielfältige Kontrollmöglichkeiten. Die Spielbarkeit ist zentral für den Erfolg eines Games. Ein Spiel, sei es sonst noch so gut, technisch wie auch vom Spielprinzip her, leidet erheblich unter schlechter Bedienbarkeit. Auch hier bieten die neuen Geräte teilweise Lösungen: So besitzen viele Mobiltelefone eine Art Steuerkreuz (z. B. Nokia 8800) oder einen Joystick (z. B. Nokia 7710 Multimedia Smartphone) [37], wodurch die Bedienung von Games erheblich erleichtert wird. Die Grösse der Tastenfelder oder Joysticks wird sich wie bei den Displays nicht wesentlich ändern. Dies fordert wiederum die Entwickler Spiele zu entwerfen, die mit den beschränkten Möglichkeiten und Kontrollmechanismen mobiler Geräte auskommen: Schnelle und komplizierte Tastaturkombinationen, die zum Beispiel bei Beat‘em-Up Games oder bei Echtzeitstrategiespielen notwendig sind, eignen sich nicht für mobile Spiele.

### Begrenzte Speicherkapazität

Während ID-Softwares DOOM3 2.2 Gigabyte Festplattenspeicher und ein Minimum von 384 Megabyte Arbeitsspeicher auf einem PC verlangt [9], besitzen die meisten mobilen Geräte wenige Kilobyte Speicher - typischerweise zwischen 128 - 500 Kilobyte. Neue Gräte können zwar schon einige Megabyte an internem Speicher anbieten (z. B. Nokia 8800 64 Megabyte interner Flashspeicher) [37] oder es ist möglich, den Speicherplatz durch zusätzliche aber teure Speicherkarten zu erweitern. Für die Spielentwickler bedeutet dies, dass Games so klein wie möglich gehalten werden müssen, indem auf spektakuläre Grafikeffekte und komplexe Sounds verzichtet wird.

### Prozessorleistung

Grafisch anspruchsvolle Spiele, wie das oben erwähnte DOOM3 stellen enorme Anforderungen an die Rechenleistung eines Gerätes und erfordern neben der CPU spezielle hoch entwickelte Grafikchips. Solche Präsentationen, wie sie heute bei Spielkonsolen oder im PC-Bereich Standard sind, können auf mobilen Geräten nicht erreicht werden. Dies wird

sich natürlich mit der technologischen Entwicklung ändern, und auch kleine mobile Geräte werden in der Lage sein, immer komplexere Grafiken darzustellen. Für Symbian OS (vgl. 7.5.1) existieren von Drittanbietern 3D-Engines. Die Applikationsplattform BREW von Qualcomm (vgl. 7.5.2) residiert direkt auf der Hardware der Endgeräte und erlaubt mit C/C++ schnelle native Anwendungen und 3D-Engines zu entwickeln. J2ME (vgl. 7.5.3), das von den meisten neuen Geräten unterstützt wird, erlaubt eine Programmierung von komplexen User-Interfaces, Multimedia Inhalten und 3D-Grafiken. Abbildung 7.5 zeigt einen Vergleich des PC-Games Splinter Cell Chaos Theory und eine Umsetzung für Mobiltelefone.


Abbildung 7.5: Vergleich Umsetzung von Splinter Cell auf PC und Mobiltelefonen Ubisoft [60], [13]

### Stromversorgung

Auf Grund der Mobilität muss ein Gerät von einer internen Batterie mit Strom versorgt werden. Diese limitierte Energieversorgung stellt sowohl für die Hardware wie auch für die Spiele selbst eine Einschränkung dar: Grössere Rechenleistung, bessere Displays und Soundausgabe erfordern bedeutend mehr Strom. Für Spiele bedeutet dies, dass nur eine beschränkte Spielzeit zur Verfügung steht, da besonders solche Geräte auch für andere Zwecke als nur zum Spielen gebraucht werden. Die Entwickler müssen daher Spiele anbieten, die auch in kurzen Spielesessionen Spass machen. Die Gerätehersteller hingegen müssen ein effizientes Energiemanagement implementieren.

### Vielfalt der mobilen Endgeräte und Portabilität

Ein weiteres Problem für die Enzwicklung mobiler Spiele stellt die enorme Vielfalt an mobilen Endgeräten dar. Dafür gibt es mehrere Gründe [27]:

- Besucht man die Websites bekannter Hersteller von Mobiltelefonen sieht man, dass jeder eine ganze Reihe von Modellen anbietet: So finden sich allein bei Nokia 52 mehr oder weniger aktuelle Modelle [37]. Samsung hat 14 aktuelle Modelle und mehrere

ältere Modelle im Angebot [46]. Und bei Sony Ericsson findet man 26 verschiedene Modelle [52]. Jedes Modell ist für einen bestimmtem Zweck entworfen und richtet sich an eine spezielle Zielgruppe: Jugendliche, Geschäftsleute, preisbewusste Benutzer, Technik-Freaks etc. Ein Mobiltelefon ist ein persönlicher Gegenstand und jeder will sein eigenes Modell, das zu ihm passt. Dies zwingt Hersteller dazu viele Modelle anzubieten, um sämtliche Kunden mit ihren Wünschen anzusprechen.

- Hersteller müssen sich von Produkten anderer Hersteller unterscheiden, um bessere Chancen am Absatzmarkt zu haben. Dies führt dazu, dass die Hersteller Modelle anbieten, die sich in vielen Punkten unterscheiden (Display, Speicher, Funktionen, Betriebssysteme, User-Interface etc.).

- Provider versuchen sich ebenfalls untereinander abzugrenzen, indem sie verschiedene Services anbieten und zu diesem Zweck in Zusammenarbeit mit Herstellern die Geräte individuell gestalten.

Es leuchtet ein, dass diese Vielfalt an Modellen ein grosses Problem für die Entwickler von mobilen Spielen mit sich bringt. Damit ein Spiel finanziell erfolgreich ist, muss dieses auf verschiedenen Modellen lauffähig sein, um so eine genügend grosse Anzahl an potentiellen Kunden zu erreichen. Dieser Portierungsaufwand für verschiedene Geräte ist extrem aufwändig. Selbst J2ME mit der Philosophie, "Write once, run anywhere.", löst dieses Problem nur ansatzweise: Geräte unterstützen unterschiedliche Standards von J2ME und bieten verschiedene zusätzliche API-Packages an. Auch die Implementierung der JRE unterscheidet sich zwischen Geräten; so zum Beispiel im Bezug auf Thread- und Speichermanagement. Die Eigenheiten und Bugs der einzelnen Modelle zu kennen und zu bewältigen, ist mit viel Aufwand und technischem Know-how verbunden. Neben den Unterschieden in der Software gibt es die schon mehrfach erwähnten Differenzen in der Hardware: Ein Spiel für verschieden grosse Displays, Prozessoren etc. anzupassen, erfordert viel Arbeit und Testen. So hat die heutige J2ME Spielindustrie originale Entwickler, die eine generische Form des Spiels entwickeln. Bei Erfolg werden auf Portierung spezialisierte Experten mit der Umsetzung des Spiels für andere Modelle beauftragt [27]. Je weniger leistungsfähig ein Modell ist, desto schwieriger ist eine solche Portierung. In der Regel sind aber gerade solche low-end Geräte weiter verbreitet als high-end Geräte.

Abildungen 7.6 zeigt, wie unterschiedlich die Spiele auf verschiedenen Geräten aussehen können (Beispiel: Splinter Cell Chaos Theory):

### Netzwerkkommunikation

Bei Mobiltelefonen und Smartphones ist die (kabellose) Netzwerkfähigkeit immanent. Dies legt nahe, Spiele zu entwickeln, die diese Kommunikationsfähigkeit ausnutzen, zu mal im PC-Bereich Multiplayer-Games in Netzwerken und im Internet immer beliebter werden. Bei mobilen Geräten ist dies mit Problemen verbunden: Die Synchronisation von mehreren mobilen Geräten der jeweiligen Spieler stellt eine Herausforderung dar. Eine weitere

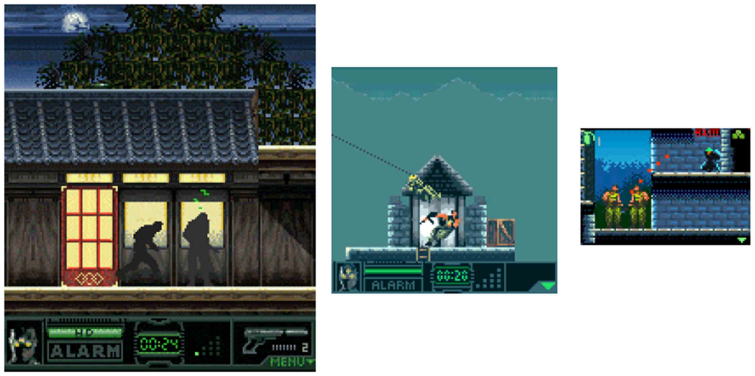


Abbildung 7.6: Unterschiedliche Präsentation auf mobilen Geräten [13]

Schwierigkeit ist die Latenz, die in mobilen Netzwerken um ein vielfaches höher ist als bei Spielen im Internet [2].

### 7.4.2 Nicht technisch bedingte Eigenschaften

Neben den technischen Eigenschaften, die Mobilität mit sich bringt, gibt es eine Reihe weiterer Punkte, die bei der Entwicklung von mobilen Spielen eine Rolle spielen.

#### Kurzweiligkeit

Mobile Spiele sind in der Regel eine Nebenbeschäftigung. Sie werden mehrmals aber nur für kurze Zeit gespielt: In öffentlichen Verkehrsmitteln, in Warteschlangen, kurz vor dem Einschlafen etc. Ein Mobiltelefon wird früher oder später auch für Anrufe verwendet. Ein Benutzer, wird also darauf achten, dass er nicht mit Spielen die ganze Batterie verbraucht und dann nicht mehr erreichbar ist. Die Spielprinzipien müssen daher so ausgelegt sein, dass sie immer wieder in kurzen Spielesessionen Spass machen, unabhängig von Zeit und Ort.

#### Abhängigkeit von Mobilfunkanbieter

Der Markt von mobilen Spielen wird durch die Mobilfunkanbieter kontrolliert. Der Anbieter bestimmt die Dienste, welche er seinen Kunden anbietet und welche Gerätefunktionen er unterstützt. So zum Beispiel, ob es möglich ist neue Spiele über das Mobilfunknetz auf das Mobiltelefon zu laden. Der Anbieter bestimmt den Preis für die Daten, die über sein Netz übertragen werden. Er übernimmt des weiteren Rechnungsstellung und hat direkten Kontakt zu seinen Kunden. Dies erleichtert das Marketing und den Support, falls Probleme auftreten. Für Entwickler ist daher wichtig, dass sie so früh wie möglich mit Mobilfunkanbietern zusammen arbeiten, da 90% des aktuellen Umsatzes über deren Portale generiert wird [27].

[2] Die Latenz im kabelgebundenen Internet wird in Millisekunden gemessen, in mobilen Netzwerken hingegen in Sekunden [35]

**Kopierschutz**

Für die meisten mobilen Spiele wird einmal beim Download bezahlt. Für den Entwickler bedeutet dies, dass ein Spiel nur bei einer genügend grossen Anzahl von Downloads rentiert. Es ist daher wichtig, dass Spiele nicht beliebig zwischen den Geräten kopiert werden können, sondern nur von demjenigen Benutzer gespielt werden können, der dafür bezahlt hat.

**Qualität**

Obwohl mobile Spiele PC- oder Konsolenspielen technisch gesehen unterlegen, bei weiten nicht so teuer und in der Regel nur ein kurzer Zeitvertreib sind, gilt auch hier das Prinzip: "Qualitiy is king". Es ist daher notwendig einen Art Qualitätskontrolle einzuführen, welche einen gewissen Standard garantiert und verhindert, dass der Markt mit schlechter Qualität überflutet wird.

### 7.4.3 Möglichkeiten und Vorteile mobiler Spiele

Neben den genannten Problemen bieten mobile Spiele enorme Möglichkeiten und Vorteile, die bei den Marktteilnehmern grosse Hoffnungen wecken:

- **Marktgrösse** (vgl. 7.1, 7.3.1): Ein gewichtiger Vorteil ist die Grösse des Marktes und die bereits vorhandene Basis an Endgeräten bei den Benutzern. Mehr als 1 Milliarde Mobiltelefone waren im Dezember 2002 in Betrieb [37]. Damit bilden mobile Netzwerke die grösste Computerplattform.
- **Neue Einnahmequellen**: Sowohl für die Netzbetreiber als auch die Entwickler bieten mobile Spiele eine neue Einnahmequelle. Neben den Sprachdiensten versuchen Netzbetreiber immer wieder neue Datendienste auf ihren Netzen anzubieten; mit bescheidenem Erfolg - eine Ausnahme bildet SMS. So sind gerade Multiplayer-Spiele sehr interessant, da sie einen permanenten Datenverkehr erfordern. Ein aktuelles Beispiel ist das bereits in Kap. 7.2.1 Spiel World of Warcraft bei dem zusätzlich Kosten für die Onlineplattform anfallen. Die Verkaufzahlen im traditionelle PC- und Konsolensektor stagnieren und die Entwickler und Publisher sehen sich einem zunehmenden Wettbewerb ausgesetzt [15]. Hier bietet sich die Möglichkeit für eine Neulancierung bekannter Spiele vom PC- und Konsolenmarkt. Es fallen hierbei nur Portierungskosten an [3]. Das bereits stehende Spielprinzip und der Bekanntheitsgrad können genutzt werden.
- **Mobilität**: Die Mobilität bringt viele Einschränkungen mit sich, ist aber ein entscheidender Vorteil gegenüber den PC- und Konsolenspielen. Auch wenn Mobiletelefone nicht primär zum Spielen entwickelt werden, tragen die Benutzer ihre Telefone immer mit sich. Dies gibt ihnen die Freiheit zu spielen, wo sie wollen und wann sie

[3] Die Portierung kann jedoch sehr aufwändig sein (vgl. 7.4.1)

wollen. Nintendos GameBoy ist nicht umsonst das Gameset, das sich am längsten im Markt gehalten hat und von dem am meisten Einheiten verkauft wurden.

- **Kosten**: Aktuelle internationale Spiele haben Budgets zwischen 2 und 20 Millionen Euro [1]. Besonders heutige Grafikstandards machen einen grossen Teil der Entwicklungskosten aus: In aufwändigen Prozessen muss eine Grafikengine entwickelt oder gegen teure Gebühren eine solche lizenziert werden. Die beschränkten Fähigkeiten mobiler Geräte dagegen erlauben bereits Entwicklungen von Spielen mit Budgets um 100‘000 Dollar. Damit sind die Einstiegsbarrieren für neue Entwicklerstudios bei mobilen Spielen viel kleiner. Das geringere finanzielle Risiko erlaubt Entwicklern auch neue Ideen auszuprobieren.

- **Entwicklungszyklus**: Der Entwicklungszyklus von mobilen Spielen ist mit einigen Monaten [35] viel kleiner (z.B. HalfLife2 5 Jahre [45]) und erlaubt einen viel kürzeren Time-to-Market Prozess und damit eine schnellere Einnahmenrealisierung.

- **Offene Standards**: Im Gegensatz zu Konsolenspielen, für die Abgaben und die Erlaubnis der betreffenden Konsolenfabrikanten notwendig sind, fallen keine Abgaben an die Hersteller von Mobiltelefonen an. Viele Standards für die Entwicklung mobiler Spiele sind offen, zum Beispiel J2ME von Sun.

Am Schluss soll erwähnt werden, dass Spiele wie Pokemon der Firma Nintendo oder Tetris beweisen, dass keine aufwändige Präsentation notwendig ist, um erfolgreich zu sein, wenn ein durchdachtes Spielprinzip vorliegt. In diesem Sinne können die technischen Grenzen mobiler Geräte auch positiv sein, indem vermehrt wieder das Spielprinzip im Mittelpunkt steht.

### 7.4.4 Kriterienkatalog

In Abschnitt 7.4 wurde versucht, mobile Spiele aus verschiedenen Blickwinkeln zu betrachten. Mit Hilfe dieser Analyse kann nun ein Kriterienkatalog erstellt werden. Dieser Katalog soll als Ausgangspunkt dienen, um im nächsten Kapitel mögliche Technologien für mobile Spiele auf deren Eignung zu untersuchen. Der Katalog wurde so gestaltet, dass die Faktoren berücksichtigt werden, die den Erfolg mobiler Spiele massgeblich beeinflussen:

- **Präsentation**: Welche Möglichkeiten bietet eine Technologie, um ein Spiel zu gestalten? Hier zählen vor allem:
  - Grafik
  - Multimedia
  - User-Interface

- **Portabilität**: Wie einfach ist es, plattform- und geräteübergreifende Spiele zu entwickeln? Wie einfach ist die Portierung auf andere Geräte und Plattformen?

- **Hardwareanforderungen**: Da viele low-end Geräte im Gebrauch sind, ist es wichtig, dass nicht allzu grosse Anforderungen an die Leistung der Geräte gestellt werden.

- **Offener oder proprietärer Standard**: Um die Entwicklung zu vereinfachen, ist ein offener Standard notwendig, an dem alle teilnehmen können.
- **Native API**: Native Programmierung erlaubt die Entwicklung schneller Anwendungen wie zum Beispiel 3D-Grafikengines.
- **Marktanteil**: Je grösser der Marktanteil, desto grösser ist die potentielle Anzahl an Geräten, auf denen die Spiele installiert werden können.
- **Distribution**: Wie einfach ist es Spiele auf den Endgeräten zu installieren? Ist Over-the-Air Distribution möglich? Wie findet die Abrechnung der Gebühren statt?
- **Kopierschutz und DRM**: Ist ein Mechanismus möglich, der die illegale Verbreitung von Spielen unterbindet?
- **Qualitätskontrolle**: Gibt es eine Möglichkeit die Qualität der Spiele zu gewährleisten?
- **Dienstleistungen**: Gibt es zusätzliche, vom Hersteller einer Technologie, angebotene Dienstleistungen, welche die Entwicklung und Vermarktung mobiler Spiele erleichtern?
- **Netzwerk**: Ist Netzwerkunterstützung für Multiplayer-Spiele, High-Score Boards etc. vorhanden?

## 7.5 Technischer Überblick

Dieses Kapitel stellt in einem ersten Schritt die wichtigsten Betriebsysteme und Applikationsplattformen für mobile Geräte vor. Diese werden anschliessend an Hand der Kriterien aus Kapitel 7.4 in ihrer Eignung für mobile Spiele eingestuft.

### 7.5.1 Symbian

Symbian wurde 1998 als private, unabhängige Firma von Ericsson, Nokia, Motorola und Psion gegründet. Symbian entwickelt und lizenziert mit Symbian OS das am weitesten verbreitete Betriebssystem für Mobiltelefone. Allein 2004 wurden 14.4 Millionen auf Symbian OS basierende Mobiltelefone versandt. Gegenüber 2003 mit 6.7 Millionen Geräten ist dies eine Steigerung von 116% [58]. Symbian befindet sich heute im Besitz von Sony-Ericsson, Ericsson, Samsung, Siemens, Nokia und Panasonic. Abbildung 7.7 zeigt die Anteile der jeweiligen Firmen an Symbian.

Dabei wird innerhalb der Firma klar zwischen Eigentümern und Management unterschieden. Technologische Angelegenheiten werden nur innerhalb des Managements behandelt. Die Eigentümer setzen lediglich die Rahmenbedingungen für Symbian OS. Symbian hat sich zum Ziel gesetzt, die Hersteller von mobilen Geräten in einem offenen Standard zu

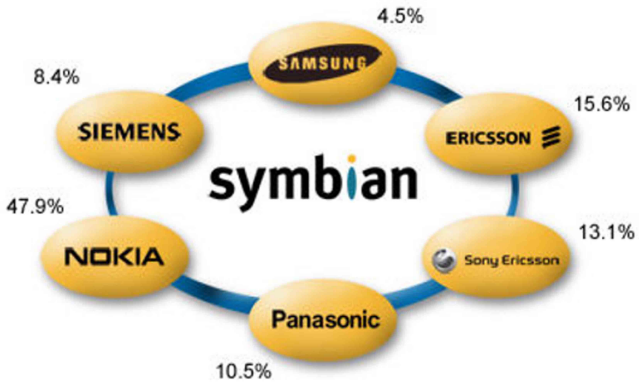


Abbildung 7.7: Firmenanteile an Symbian [58]

vereinen. Symbian versucht diese Communitiy-Strategie über Partnerprogramme zu verstärken: Platinum Partner Program und Affiliate Partner Program. Bei diesen Programmen können alle teilnehmen, die in irgendeiner Weise mit Symbian in Verbindung stehen: Mobilfunkanbieter, Hardwarehersteller, Applikationsentwickler, Consultingfirmen etc. Je nach Art des Programms können die Mitglieder von verschiedenen Vorteilen profitieren: Technischer Support, exklusive Voreinsicht in den Sourcecode von Symbian OS, Zugang zu Vertriebskanälen, Marketingevents, Conventions, zu einem Netzwerk von Marktteilnehmern, Möglichkeit sich an der Entwicklung von Symbian OS zu beteiligen etc. Diese Strategie soll Herstellern von Mobiltelefonen die Möglichkeit bieten, auf der einen Seite differenzierte Modelle anzubieten, aber auf der anderen Seite eine gemeinsame Plattform zu nutzen.

Symbian OS versteht sich nicht als Weiterentwicklung bestehender Technologien, sondern als zukunftsgerichtet disruptive Technologie, welche die alten "embedded" Betriebssysteme ersetzt. Symbian OS wurde speziell für die Anforderungen moderner Mobiltelefone entwickelt. Die neuste Version Symbian OSv9 erschien im Februar 2005. Version 9 richtet sich daher an den Massenmarkt mit Geräten der 2.5G und 3G [58]. Symbian soll helfen, die Kosten und den Time-to-Market Zyklus bei der Entwicklung moderner Mobiltelefone zu reduzieren. Abbildung 7.8 zeigt die aktuellen Lizenznehmer von Symbian OS.

### Schlüsselelemente von Symbian OSv9.1

Die folgende Liste ist nicht abschliessend. Eine vollständige Beschreibung und Spezifikation von Symbian OSv9.1 und früheren Versionen finden sich der Website von Symbian: http://www.symbian.com

- **Unterstützung von Java**: Unter anderem mit CLCD 1, MIDP 2.0, Mobile 3D Graphics API
- **Multimedia**: Audio- und Videoaufnahme, -wiedergabe, -streaming
- **Grafik**:
  - Erhöhte Flexibilität bei User-Interfaces (z.B. mehrere simultane Displays)

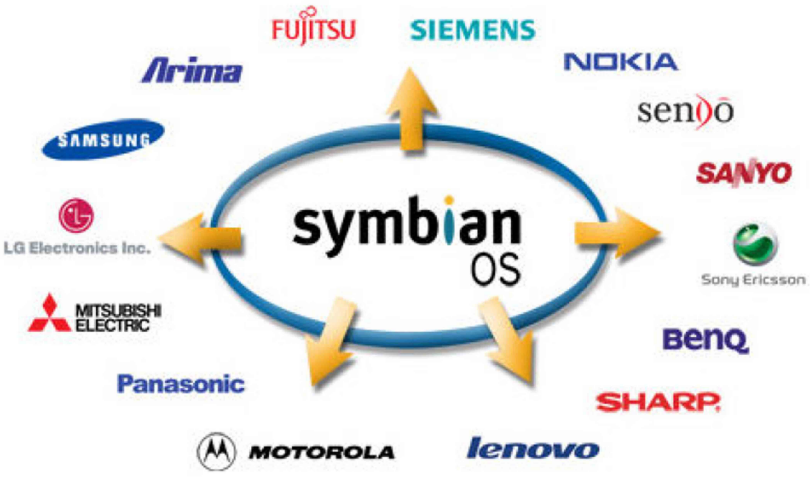


Abbildung 7.8: Lizenznehmer von Symbian OS [58]

- Unterstützung der Khronos 3D OpenGL ES API[4]
- Direkter Hardwarezugriff auf Display und Keyboard

Abbildung 7.9 zeigt auf Symbian OS basierende User-Interfaces:

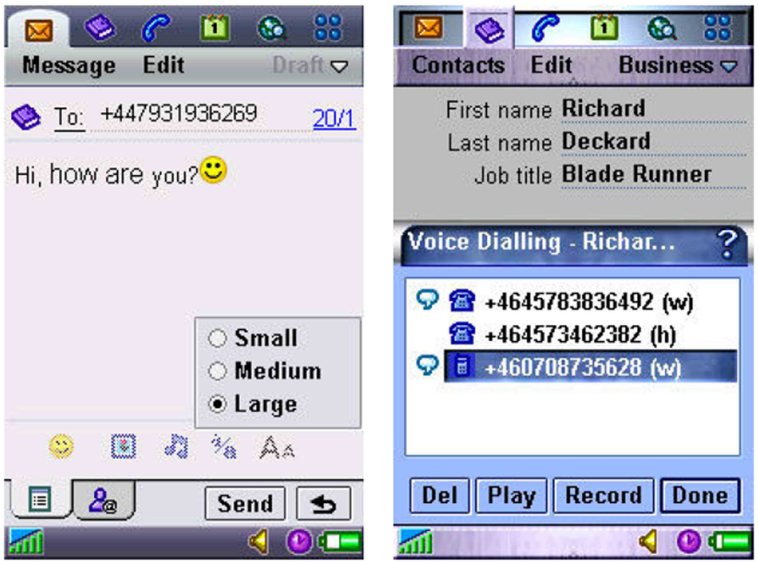


Abbildung 7.9: Symbian OS basierte User-Interfaces [58]

- **Realtime und Hardwareunterstützung**:
  - Realtime, multithreaded Kernel
  - Unterstützung neuer CPU-Architekturen (z.B. ARM Prozessoren[5] und Speichermodule)

[4]OpenGL ES ist eine kleine low-level API für plattformübergreifende Grafikprogrammierung auf mobilen Systemen. OpenGL ES basiert auf der OpenGL Technologie [24]

[5]ARM ist ein Technologiekonzern, der unter anderem Prozessoren mit geringem Stromverbrauch herstellt [3]

  - Effizientes Energiemanagement

- **Netzwerk- und Nachrichtenkommunikation**:
  - TCP/IP (IPv4/IPv6 Dual Mode)
  - Personal Area Networking (Infrarot, Bluetooth, USB)
  - E-Mail (POP3, IMAP4, MIME Attachments)
  - Enhanched Messaging (EMS)[6]
- **Over-the-Air (OTA)**:
  - OTA Synchronisation von Daten (z. B. für Agenda) basierend auf Open Mobile Alliance Standards
  - OTA Gerätemanagement (Updates, Programme etc.)
- **Mobiltelefonie**: GSM, GPRS, HSCSD, CDMA
- **Sicherheit**:
  - DRM-Framework
  - IPSec und VPN Client Unterstützung
  - Zertifikate (z. B. Symbian Signed)
- **Programmiersprachen**:
  - Native Programmierung mit C++
  - Java
  - OPL (Open Programming Language)
  - Microsoft .NET [7]

Abblidung 7.10 zeigt das Systemmodell von Symbian OSv9.1

### 7.5.2 BREW

BREW ist eine von Qualcomm entwickelte offene, standardisierte und erweiterbare Applikationsplattform für mobile Geräte. BREW ist vor allem in den USA verbreitet.

Der Hauptanreiz von BREW liegt in der End-to-End Philosophie. Qualcomm versucht mit BREW die ganze Wertschöpfungskette zu vereinigen. Für jedes einzelne Glied der Kette (Hersteller von mobilen Geräten, Entwickler, Publisher, Content Provider, Netzbetreiber, Endkunden) bietet BREW spezielle auf dessen Bedürfnisse abgestimmte Dienste und Unterstützung an. Die BREW Applikationsplattform selbst ist daher nur ein Teil von The BREW Ecosystem [5].

[6] EMS erweitert SMS um Bilder, Töne und Animationen. EMS ist ein Zwischenschritt hin zu MMS.
[7] Für .Net wird allerdings das Tool Crossfire von AppForge benötigt [58].

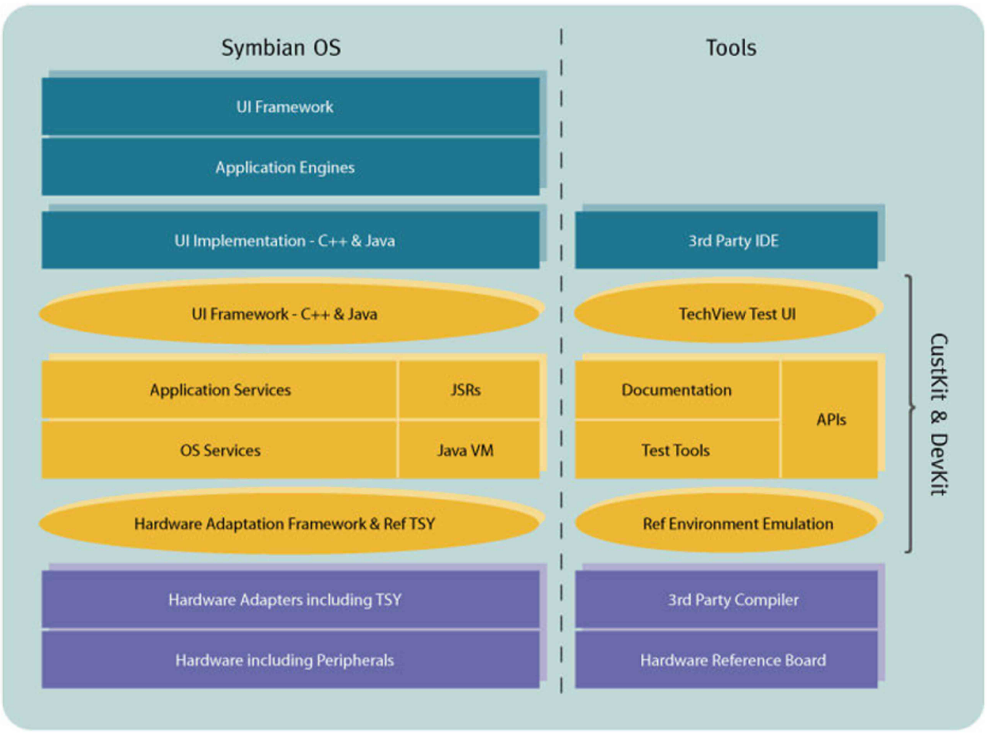


Abbildung 7.10: Systemmodell von Symbian OSv9.1 [58]

### The BREW Ecosystem

- **BREW Applications Platform**: Die Applikationsplattform des BREW System befindet sich direkt auf der Hardware des Gerätes. Dies ermöglicht schnelle native Applikationen in C/C++ und erlaubt eine einfache Integration von Java, sowie eine Reihe von Erweiterungen zum Beispiel 3D-Engines, XML-Parser, Videoplayer. BREW ist ein offener erweiterbarer Standard, der unabhängig von Betriebssystem auf einem Gerät eingesetzt werden kann. BREW beinhaltet eine Standardmenge von API, die auf allen Geräten konsistent ist. Dies macht es möglich, Applikationen ohne grossen Aufwand auf verschiedene Geräte zu portieren. Sowohl Geräte- wie Dritthersteller können Erweiterungen und Applikationen für BREW entwickeln, die Over-the-Air vertrieben werden (erlaubt zum Beispiel bequemes OTA Bug-Fixing). BREW unterstützt verschiedene Authentifizierungsverfahren und verwendet digitale Signaturen um Transaktionen und OTA Downloads zu verifizieren. Mit den BREW Porting Kit kann die BREW Applikationsplattform auf einfache Art und Weise in mobile Geräte integriert werden.

- **BREW SDK**: Das BREW SDK bietet Entwicklungstools, um Applikationen zu entwickeln. Dabei ist keine Kenntnis über die Hardware des Gerätes notwendig.

- **BREW Distribution System (BDS)**: Über einen virtuellen Marktplatz verbindet das BDS Publisher, Entwickler und Netzoperatoren. Das BDS ermöglicht einen sicheren OTA Vertrieb von Applikationen, koordiniert Rechnungsstellung und Bezahlung, Support und Beratung. Das BDS richtet sich in erster Linie an Netzbetreiber.

- **BREW Value Added Services**:
  - Developertools and Resources: Unterstützung für die Applikationsentwicklung: IDE, Compiler, Marketingsupport, Datenbanken etc.
  - Developer Relations Support: Hilft Vertreibern und Herstellern von Applikationen, diese auf dem Markt zu lancieren.
  - BREW Training: Workshops und Tagungen mit BREW-Experten sollen helfen, erfolgreiche Applikationen zu entwicklen.
  - BREW Product Support.
  - TRUE BREW Testing (TBT): Dient der Qualitätssicherung.
  - Application Signing: Einheitliche Überprüfung von BREW Publishern und Entwicklern. Diese werden mit einer digitalen Signatur versehen, um die Integrität der BREW Applikationen zu gewährleisten.
  - Device Manufacturer Support: Unterstützt Hersteller, um eine schnelle und konsistente Implementierung von BREW auf allen Geräten zu erreichen.

Für Entwickler mit wenig Erfahrung im mobilen Markt bietet BREW mit seinem System eine einfache Einstiegsmöglichkeit. Die starke Integration sämtlicher Beteiligter an der Wertschöpfungskette erschwert aber auch die Vermarktung von Applikationen von unabhängigen Drittherstellern, falls diese keinen Zugang zum BREW Netzwerk besitzen.

### 7.5.3 J2ME

Die Java 2 Platform Micro Edition (J2ME) ist wohl die populärste Applikationsentwicklungsplattform für mobile Geräte. J2ME ist auf den meisten Mobiltelefonen verfügbar, wie z.B. bei Nokia, Motorla, Samsung und Sony-Ericsson. Da diese Hersteller zusammen ca. 80% Marktanteil haben, ist J2ME auf weit mehr als 250 Mio. Geräten verfügbar [27].

J2ME enthält wie die anderen Umgebungen - Enterprise (J2EE), Desktop J2SE) und smart card (Java Card) - eine virtuelle Maschine und ein Set von Standard Java APIs. Zusätzlich ermöglicht sie die Programmierung von flexiblen User-Interfaces, stellt ein robustes Sicherheitsmodell und eine breite Palette von Built-in Netzwerkprotokollen zur Verfügung und gibt weitreichenden Support für netzwerkbasierte und offline Applikationen, welche dynamisch heruntergeladen werden können [53].

#### J2ME Architektur

Wie in Abbildung 7.11 ersichtlich, beinhaltet und kombiniert die J2ME Architektur eine Menge von Konfigurationen, Profilen und optionalen Packages, aus welchen Implementierer und Entwickler wählen können, um eine komplette Java Runtime Umgebung zu konstruieren. Jede Kombination ist optimiert für den Speicher, die Prozessorleistung, und die I/O Ressourcen der jeweiligen Kategorie des Gerätes. Das Resultat ist eine Java Plattform, welche den Vorteil jedes Gerätetyps benutzt und so das Maximum aus diesen herauskitzelt [53].

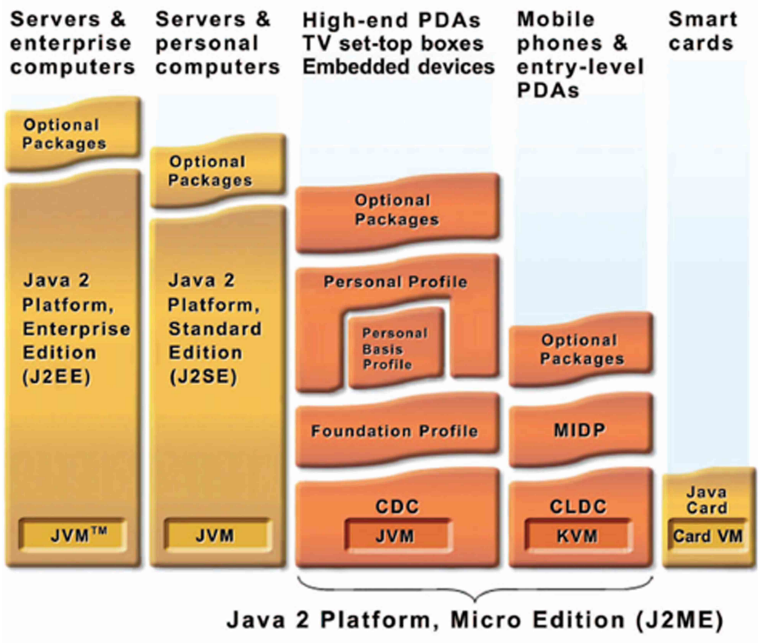


Abbildung 7.11: Eingliederung von J2ME [53]

- **Konfigurationen**: Die Konfigurationen beinhaltet eine Virtuelle Maschine und ein minimales Set von Klassenbibliotheken. Diese stellen die Grundfunktionalität für ein bestimmtes Segment von Geräten, welche ähnliche Charakteristiken, wie z.B. Netzwerkkonnektivität und Speichergrundlagen, haben, zur Verfügung. Momentan gibt es zwei J2ME Konfigurationen: Die Connected Limited Device Configuration (CLDC) und die Connected Device Configuration (CDC).

- **Profile**: Um eine komplette Runtime-Umgebung für ein spezielles mobiles Gerät zu haben, muss eine Konfiguration mit einem Profil kombiniert werden, d.h. mit einem Set von higher-level APIs. Diese definieren das Applikations Lebenszyklusmodel, das User-Interface und den Zugang zu gerätespezifischen Eigenschaften. Ein Beispiel dafür ist die Kombination von CLDC mit dem Mobile Information Device Profile (MIDP), welche eine komplette Java Applikationsumgebung für Mobiltelefone bereit stellt.

- **Optionale Packages**: Die J2ME Plattform kann mit verschiedenen Packages erweitert werden. Sie werden gebraucht, um sehr speziellen Applikationsanforderungen zu genügen, wie z.B. Datenbankverknüpfung, Wireless Messaging, Multimedia, Bluetooth und Webservices.

Im folgenden Abschnitt wird der CLDC-MIDP Teil des J2ME genauer betrachtet, da vorallem diese Kombination in Mobiltelefonen ihre Anwendung findet und für die zukünftige Entwicklung des Marktes am meisten von Bedeutung ist.

### CLDC

Die Konfiguration CLDC beinhaltet als virtuelle Maschine die K Virtual Machine (KVM). Diese ist eine hochportable virtuelle Maschine, welche besonders für netzwerkverbundene Geräte mit wenig Speicher und limitierten Ressourcen geeignet ist - wie z.B. für Mobiltelefone. Die KVM ist klein (statischer Speicher von 40-80kbytes), gut dokumentiert, hochportabel, modular, anpassbar und so komplett und schnell wie möglich, ohne dabei andere Ziele zu beeinflussen [54].

Die CLDC Konfiguration beinhaltet folgende Bereiche [54]:

- Javasprache und Merkmale der virtuellen Maschine
- Wichtigsten Java-Bibliotheken (java.lang.*;java.util.*)
- Input/Output
- Networking
- Sicherheit:
  - Low-level-virtuelle Maschine Sicherheit, indem heruntergeladene Javaklassen einen Classfile Verifikationsschritt durchlaufen müssen.
  - Applikationen sind voneinander geschützt, da sie jeweils einzeln in einer Sandbox-Umgebung laufen.
  - Klassen in geschützten Systempackages können nicht von Applikationen überschrieben werden.
- Internationalisation

### MIDP2.0

Das Profil MIDP enthält Klassen und Routinen zur Ausgabe und Verarbeitung von Grafiken, zur Verwaltung von Systemmenüs sowie zur Interaktion mit der Firmware, hierbei im speziellen Systemfunktionen wie der System Timer, File Handling, aber auch Kommunikation, Multithreading und vieles mehr. Die Implementierung der Funktionen selbst ist hierbei Sache des Geräteherstellers [59].

Eine detailliertere Übersicht über die Merkmale von MIDP liefert die Abbildung 7.12:

- **Mobiles User Interface**: MIDP enthält ein high-level User Interface API, welches den Entwickler von der Komplexität der Erstellung von portablen Applikationen abschirmt. Das User-Interface enthält vordefinierte Raster um Listen anzuzeigen und zu selektieren, Texteditiermöglichkeiten und Anzeige von Alarmnachrichten.

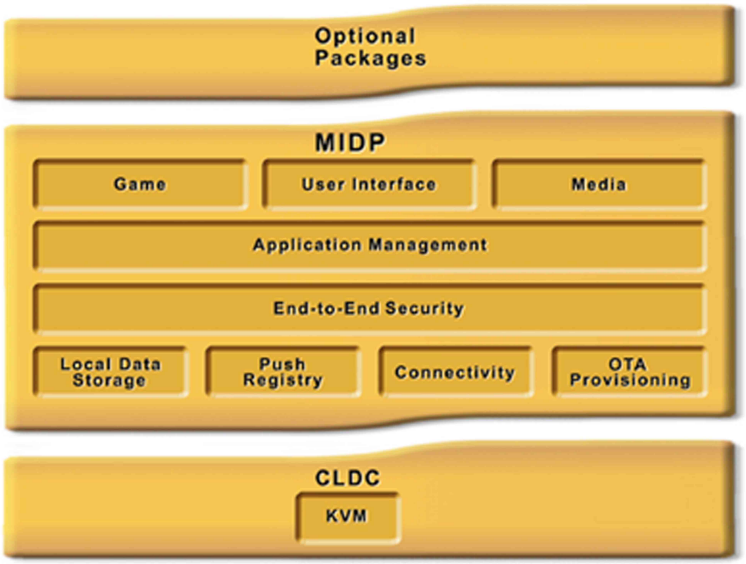


Abbildung 7.12: Übersicht über das Profil MIDP [55]

- **Multimedia and Game Functionality**: Ein Game API fügt gamespezifische Funktionalitäten, wie z.B. sprites and tiled Layers, Game Canvas und Imagehandling hinzu. Built-in Audio ermöglicht die Unterstützung von Tönen, Tonsequencen und Wav-Files. Zusätzlich kann die Mobile Media API (MMAPI) benutzt werden, um Videos oder andere Multimediainhalte einzubinden.
- **Umfassende Konnektivität**: MIDP ermöglicht den Zugriff und die Benutzung des Datennetzwerk und Messagingmöglichkeiten der mobilen Geräte. Es unterstützt einerseits HTTP und HTTPS, Datagram, Sockets, Server Sockets und Serial Port Kommunikation. Durch das Wireless Messaging API (WMA) Package unterstützt MIDP auch die SMS Fähigkeiten von GSM and CDMA Netzwerken.
- **Over-the-Air Provisioning**: Applikationen können over-the-Air (OTA) verteilt und aktualisiert werden. Die MIDP-Spezifikationen zeigt, wie Applikationen installiert, aktualisiert und gelöscht werden. Über MIDP kann ein Service Provider identifizieren, welche Applikationen auf einem bestimmten Gerät laufen, und den Status während einer Installation überprüfen.
- **End-to-End Security**: MIDP stellt ein robustes Sicherheitsmodell zur Verfügung:
  - HTTPS ermöglicht die Übertragung von verschlüsselten Daten.
  - Sicherheits-Domains schützen vor unauthorisierten Zugriff auf Daten, Applikationen und anderen Netzwerk- und Geräteressourcen.

### 7.5.4 .NET Compact Framework

Mit dem .NET Compact Framework will Microsoft in die Entwicklung von Applikationen für PocketPCs oder andere mit Windows CE ausgerüsten mobilen Geräte einsteigen. Das

Framework ist ein Subset vom Standard .NET Framework und bietet daher analog eine Applikationsplattform mit eine Serie von Runtime-Komponenten. Der grösste Unterschied ist, wie in Abbildung 7.13 ersichtlich ist, dass das darunterliegende Betriebssystem nicht Windows NT o.ä., sondern das Windows CE Betriebssystem ist [61].

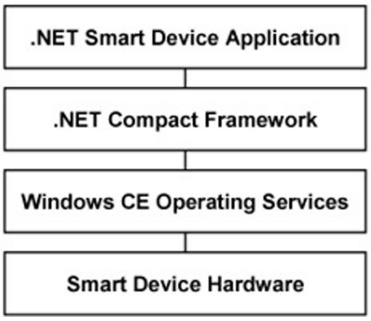


Abbildung 7.13: Position des .NET CF in mobilen Geräten [61]

Das .NET Compact Framework ermöglicht die Erstellung von Programmen, die in einer Runtime-Umgebung laufen, vor der Ausführung kompiliert werden und hohe Funktionalität mit bestmöglicher Performance erlauben. Dieses Prinzip weist Parallelen zur Java-Technologie und deren JVM auf.

### .NET Compact Framework generelles Design

Bei der Architektur des Frameworks orientiert sich Microsoft an der PC-Architektur. Ziel ist es, für die mobilen Endgeräte ähnlich wie auf dem PC eine Plattform zu kreieren, bei der Entwickler Anwendungen programmieren, ohne sich um die hardwarespezifischen Details zu kümmern. Analog zum Windows-PC will der Softwarekonzern hierzu einen Satz von APIs als Subset von .NET definieren, auf den die Programmierer zugreifen. Microsoft stellt jedoch lediglich generische, gerätespezifische Libraries zur Verfügung. Die geschriebene Applikation wird nach dem herunterladen kompiliert. Aufgabe der als Compiler fungierenden Run-Time Umgebung ist dabei nicht nur einen schnellen Code zu generieren, sondern in den Libraries die hardwarespezifischen Features herauszusuchen und in den Code einzubinden. Dies hat zum Vorteil, dass bei neuer Hardware die Applikation an sich nicht geändert werden muss, sondern nur eine Anpassung der entsprechenden Library erforderlich ist [20].

### Microsoft Game API

Das Microsoft Game API (GAPI) zielt auf high-performance Gaming und Visualisierungsapplikationen ab und enthält nur 12 Funktionen. Diese leichtgewichtige C-basierte API ist für mobile Geräte kreiert und kann in das Betriebssystem Windows CE integriert werden. Mit der GAPI können folgende Ziele erreicht werden [32]:

- Schnelle und animierte Grafiken: GAPI erlaubt den direkten Zugriff auf den Video-Frame-Buffer-Speicher des Displays. Da jedes Gerät verschieden ist, muss GAPI jedes Gerät definieren.

- Gebrauch der Hardwaretasten auf windowsbasierten mobilen Pocket PC's für die Spielsteuerung: Erlaubt der Anwendung, die Kontrolle über die Hardwaretasten zu übernehmen.
- Verstecken der Menüleiste: Das Verwalten der Menüleiste und die Wiederherstellung des vorherigen Status wird garantiert.
- Unterbindet Unterbrechungen beim normalen Gebrauch des Geräts: Das Kontrollieren des Fokuses ist möglich.
- Aufblitzenlassen des Grundlichts für Spezialeffekte: Ein simples API wird dafür bereitgestellt.
- Gebrauch des Vibrationsalarms als Rumble Pack: Noch unter Konstruktion.

### 7.5.5 Beurteilung der Technologien anhand des Kriterienkatalogs

Die Beurteilung erfolgt mittels dem in Kapitel 7.4.4 definierten Kriterienkatalog. Die Beurteilung berücksichtigt die wichtigsten Fähigkeiten und Funktionen der jeweiligen Technologien. Der Miteinbezug der vollständigen Spezifikationen würde den Rahmen dieser Beurteilung sprengen. Die vollständigen Dokumentationen und Spezifikationen finden sich in den im Literaturverzeichnis angebenen Verweisen.

| | Symbian | BREW | J2ME | .NET Compact Framework |
|---|---|---|---|---|
| **Präsentation** | Symbian unterstützt J2ME und ermöglicht die Nutzung dessen Möglichkeit zur Audio - und Videogestalltung der Spiele (vgl. 1.5.5, 1.5.3). Symbian unterstützt die Khronos 3D OpenGL ES API und ermöglicht direkten Hardwarezugriff. Kombiniert mit nativer Programmierung C++, können schnelle 3D-Engines entwickelt werden. Ein Beispiel ist die Diesel Engine von 3D Arts [7]. Das Audio- und Videoframework erlauben, Spiele mit Multimediainhalten anzureichern. | BREW erlaubt eine einfache Entwicklung und Integration von Erweiterungen, die für eine "State of the Art" Präsentation sorgen. Beispiele dafür sind: 3D-Engines, Videoplayer, Javaunterstützung. Eine Demonstration des 3D-Potentials von BREW bietet das"3D Gaming Showcase"; http://brew.qualcomm.com/brew/en/operator/demos/ | Für die grafische Unterstützung existiert die Game API, welche gamespezifische Funktionalitäten einführt. Im multimedialen Bereich werden die mobilen Spiele durch Built-In Audio und die Mobile Media API unterstützt. Die Unterstützung für das User-Interface folgt durch die Mobile User Interface API. Für weitere Details siehe Kapitel 1.5.3. | Die Game API des .Net Compact Frameworks stellt Hilfsmittel für die grafische Darstellung, die Einbindung von multimedialen Hilfsmitteln und die Benutzung des User-Interfaces zur Verfügung, vgl. Kapitel 1.5.4. |
| **Portabilität** | Symbian versucht sein Betriebssystem als Standart auf mobilen Geräten durchzusetzen. Die Unterstützung anderer offener Standards (J2ME, OpenGLES) und Benutzung verbreiteter Programmiersprachen sollen Protabilität vereinfachen. Gerade hardwarenahe Grafikprogrammierung jedoch ist sehr geräteapezifisch und erschwert Protabilität. Eine weitere Hürde stellen die verschiedenen Versionen von Symbian OS dar. Auf Grund der Weiterentwicklung wird kaum jemals der Fall eintreten, dass alle Geräte die gleiche Version implementieren. | BREW besitz eine Standardmenge von API, die auf allen Geräten konsistent ist. Der Device Manufacturer Support im BREW Value Added Services-Programm versucht, eine solche geräteübergreifende konsistente Implementierung von BREW zu erreichen. | Die Kombination von der Konfiguration (CLDC) und dem Profil (MIDP2.0) sorgt dafür, dass eine komplette Runtime-Umgebung für ein spezielles mobiles Gerät vorhanden ist. Die Entwickler müssen sich demnach nicht mehr um die gerätespezifischen Eigenschaften kümmern, und können in einem gewissen Rahmen plattformunabhängig entwickeln. Gewisse Anpassungen für die unterschiedlichen Geräte (Displaygrösse, etc. ) sind dennoch nötig, was zu unterschiedlichen Versionen des Spiels führt. | .Net Code ist unabhängig von der Konfiguration des Systems. Einmal geschriebener Code ist auf jeder Windows-CE Plattform lauffähig. Das einzige Portabilitätsproblem ist die Beschränkung auf Windows-Plattformen [40]. |
| **Hardware-anforderungen** | Symbian OS ist klar auf Mobiltelefone und deren Möglichkeiten ausgerichtet. Symbian unterstützt daher zum Beispiel mit ARM-Prozessoren, Hardwarebestandteile die speziell für mobile Geräte entwickelt wurden. | BREW ist für mobile Geräte entwickelt worden. | Die Konfiguration CLDC hat geringe Hardwareanforderungen. | Das .NET CF hat hohe Hardwareanforderungen, da es nur auf Geräten mit dem Betriebssystem Windows CE bzw. Pocket PCs und PDAs läuft. |
| **Offener oder proprietärer Standard** | Symbian selbst ist ein offener Standard. Mit der Einbindung weiterer offener Standards wird diese Tendenz verstärkt (z. B. J2ME). | Die BREW Applikationsform wird von Qualcomm als offener und erweiterbarer Standard entwickelt. | Offener Standard, da die Spezifikationen frei verfügbar sind. | Das .NET Compact Framework ist als proprietärer Standard anzusehen. |

Abbildung 7.14: Kriterienkatalog Teil 1

| | | | | |
|---|---|---|---|---|
| **Native API** | Als Betriebssystem erlaubt Symbian native Programmierung. | Native Programmierung mit C/C++ ist möglich. | Innerhalb von CLDC nicht vorhanden [27]. In der Konfiguration CDC ist es mittels JNI möglich. | Native API wird durch P/Invoke9 unterstützt [27]. |
| **Marktanteil** | Symbian ist das am weitesten verbreitete Betriebssystem für Mobiltelefone [14]. Die wichtigsten Gerätehersteller sind an Symbian beteiligt oder Lizenznehmer (vgl. 1.5.1). | BREW ist hauptsächlich in den USA verbreitet. | Der Marktanteil ist sehr hoch, da es auf den meisten Mobiltelefonen verfügbar ist (>250 Mio. Geräte). | Im Bereich der Handhelds haben Windows-CE basierende Geräte einen Marktanteil von 48% [30]. Somit ist der Markt der potentielle Kunden für mobile Spiele auf Basis des .NET Compact Framework gross. |
| **Distribution** | Symbian untersützt OTA Transaktionen basierend auf Open Mobile Alliance Standards. | BREW ermöglicht einfache OTA Transaktionen. Die starke Integration der Wertschöpfungskette bietet enge Kontakte zu Publishern und Netzoperatoren und erleichtert die Distribution von Applikationen. Das BDS koordiniert Rechnungsstellung und Bezahlung. | Der Distributionsvorgang ist relativ einfach gehalten: 1. Applikationen können Over-the-Air verteilt und aktualisiert werden. 2. Die Installation erfolgt nach OTA-Spezifikation [5]. 3. Die Abrechnung erfolgt über den Mobiltelefon-Provider | Over-the-Air Distribution wird bis anhing nicht unterstützt. Aufgrund der Geräte (PDA und Smartphones) müssen die Spiele auf den PC heruntergeladen und von dort auf dem Gerät installiert werden. Die Zahlung erfolgt meistens mittels Kreditkarten. |
| **Kopierschutz und DRM** | Symbian OSv9 implementiert ein umfangreiches Sicherheitmodell: DRM-Framework, IPSec, Kryptographie, Zertifikatmanagement etc. | Open Mobile Alliance DRM ist in Zukunft g | Spezifische Kopierschutzmechanismen für mobile Spiele werden von Sun bzw. J2ME bisher nicht explizit angeboten, sondern wird durch die Gerätehersteller und andere Serviceanbieter vorangetrieben [22]. Allgemeine Sicherheitsmechanismen sind implementiert - vgl. Kapitel 1.5.3 | Ein Kopierschutz über ein DRM für Spiele gibt es in Windows CE bisher noch nicht (nur der Schutz für multimediale Inhalte). Allgemeine Sicherheitsmechanismen zum Schutz des mobilen Geräts sind aber verfügbar. |
| **Qualitäts-kontrolle** | Die Bildung einer "Symbian-Community" und die Zertifizierung durch "Symbian Signed" erlaubt eine gewisse Überwachung der Applikationen. | BREW bietet die Möglichkeit des TRUE BREW Testing und des Application Signing zur Sicherung der Qualität und Integrität der Applikationen. | Die Qualitätskontrolle erfolgt durch die Verleihung von Java Powered Logos ("Java Verified") und der Veröffentlichung des Spiels auf dem Java-Portal www.java.com [48]. | Nur Spiele, welche ausgiebig von den Partner von Microsoft getestet werden, bekommen das Microsoft-Zertifikat und können danach auf dem Mobile2Market's mobilen Applikationskatalog vermarktet werden [32]. |

Abbildung 7.15: Kriterienkatalog Teil 2

| | | | | |
|---|---|---|---|---|
| **Dienst-leistungen** | Mitglieder des Partnerprogramms profitieren von diversen Zusatzleistungen. Sybmian bietet auf seiner Website http://www.symbian.com Hilfe für Entwicklung von Applikationen an: Dokumentationen, Beispielcode, Entwicklertools, SDK's etc. | Der grosse Vorteil BREW liegt in der Integration und starken Bindung sämtlicher Marktteilnehmer. Dies ermöglicht BREW gezielte Unterstützung für die einzelnen Teilnehmer anzubieten. | Sun bietet verschiedene Dienstleistungen an: 1.Software und Tools: z.B. SDKs und integrierte Entwicklungsumgebungen speziell für mobile Anwendungen. 2. Webbasiertes technisches Training. 3. Go-to-Market-Services: Mittels dem "Java Device Test Suite" und Veröffentlichung im Sun Content Katalog. 4. Viele technische Dokumente, wie Spezifikationen, White Papers, Code Samples, etc. 5. Blueprints und Best Practices Dokumente. 6. Support Optionen: Durch die weltweit verfügbaren Sun Support Zentren. | Microsoft stellt weitreichende Hilfsmittel zur Entwicklung von mobilen Spielen zur Verfügung [61]: 1. Entwicklungstools: Eine Reihe von kostenpflichtigen Entwicklungstools werden zur Verfügung gestellt (z.B. Microsoft Visual Studio .NET 2003). 2. Zertifikations- und Marketingprogramme: Über das Programm Mobile2Market können unabhängige Softwareentwickler die Time-to-Market verkürzen und Zugang zu weltweiten Verteilungszentren bekommen. 3. Zugriff auf das MSDN Mobile and Embedded Developer Center. 4.Chats und Newsgroups: Zum Austausch mit anderen Entwicklern. 5. Support-Dienste: Weltweiter Support-Dienst von Microsoft. |
| **Netzwerk** | Symbian OS unterstützt umfangreiche Netzwerkkommunikation (vgl. 1.5.1). | BREW besitzt eine Netzwerk API zum Aufbauen von TCP(UDP)/IPVerbindungen. | Weitreichende Netzwerkunterstützung (HTTP, | .NET CF unterstützt allgemeine Netzwerkprotokolle und verbindet nahtlos XML-Services. |

Abbildung 7.16: Kriterienkatalog Teil 3

## 7.6 Schlussfolgerungen

In den vorhergehenden Kapiteln wurde ein Überblick über das Thema der mobilen Spiele und die aktuellen Technologien gegeben. Wie ist die aktuelle Situation zu beurteilen? Der technische Forschritt bei den Geräten erlaubt, immer komplexere Spiele zu entwicklen. Das Marktpotential ist auf Grund der grossen Anzahl vorhandener Geräte vorhanden. Länder wie Japan demonstrieren, dass ein erfolgreicher Markt mit mobilen Spielen möglich ist (vgl. Kap. 7.3.2). Auch die Beurteilung der Technologien an Hand des Kriterienkatalogs in Kapitel 7.5.5 zeigt, dass die verschiedenen Technologien durchaus in der Lage sind, die Voraussetzungen für mobile Spiele zu erfüllen. Die Bestrebungen einen gemeinsamen Standard wie beispielsweise Symbian (vgl. Kap. 7.5.1) oder J2ME (vgl. Kap. 7.5.3) zu finden, erleichtern die Entwicklung von mobilen Spielen.

Entscheidend für den Erfolg wird sein, dass die Industrie nicht den Anspruch hat eine Alternative zu traditionellen Spielen zu sein, sondern eine andere Form des Spielens - das mobile Spielen. Daher ist es wichtig, dass man versucht, Spiele und Dienste zu entwicklen, welche die Eigenschaften mobiler Spiele berücksichtigen und darauf abgestimmt sind. Gelingt dies, können mobiler Spiele als eigene Form des Spielens existieren ohne andauernd im Vergleich mit PC- und Konsolenspielen unterliegen zu müssen, und eine Etablierung im Massenmarkt ist durchaus realistisch.

# Literaturverzeichnis


[1] A. Donath,"Prognose: Mobile Spiele werden ein Milliardenmarkt in Europa", (Golem), [online] 2003,
http://www.golem.de/0311/28382.html (Accessed: 3.Mai 2005)

[2] AreaMobile News, AreaMobile, [online] 2004,
http://www.areamobile.de/news/2534.html (Accessed: 3.Mai 2005)

[3] ARM-Overview, ARM, [online] 2005,
http://www.arm.com (Accessed: 3.Mai 2005)

[4] BenQ News, BenQ, [online] 2005,
http://www.benq.de/news/index.cfm?id=995&year=2005 (Accessed: 30.Juni 2005)

[5] BREW-Overview, QUALCOMM BREW, [online] 2005,
http://brew.qualcomm.com/brew/en/ (Accessed: 3.Mai 2005)

[6] C. Enrique Oritz, "Introduction to OTA Application Provisioning", (Sun), [online] 2002,
http://developers.sun.com/techtopics/mobility/midp/articles/ota/ (Accessed: 3.Mai 2005)

[7] Das Unternehmen Sun Microsystems, Sun, [online] 2005
http://de.sun.com/company/sun/ (Accessed: 3.Mai 2005)

[8] DieselEngine® SDK, 3D Arts, [online] 2005,
http://www.3darts.fi/mobile/dieselengine.htm (Accessed: 3.Mai 2005)

[9] Doom3, System Requirements Gamecase Doom3, ID-Software

[10] Elkware Overview, Elkware, [online] 2005,
http://www.elkware.com/ (Accessed: 3.Mai 2005)

[11] Forum Palm, Metamac, [online] 2004,
http://www.metamac.de/forum/viewtopic.php?pid=1522 (Accesed: 3.Mai 2005)

[12] Gameboy Overview, Nintendo, [online] 2005,
http://ms.nintendo-europe.com/gameboyadvance/ (Accessed: 3.Mai 2005)

[13] Gameloft Overview, Gameloft, [online] 2005,
http://www.gameloft.com/corpo_company_profile.php (Accessed: 3.Mai 2005)

[14] H. Blankenstein, "Mobile Media: Von Japan lernen", (BLM), [online] 2004,
http://www.blm.de/apps/documentbase/data/de/28-29_japan.pdf (Accesed: 3.Mai 2005)

[15] H. Sidow; G. Goeltzer, "Mobile Gaming: Realisable opportunity or market hype", (Detecon), [online] 2004,
http://www.competence-site.de/mbusiness.nsf/B0F018F1AC2ABCB9C1256E960047B327/$File/mobile_gaming_detecon.pdf (Accessed: 3.Mai 2005)

[16] Handymarkt in Europa: Nokia unter Druck, Inside-Handy, [online] 2004,
http://www.inside-handy.de/news/2171.html (Accessed: 3.Mai 2005)

[17] History, Symbian, [online] 2004,
http://www.symbian.com/about/history.html (Accessed: 3.Mai 2005)

[18] I-mode, Wikipedia, [online] 2005,
http://de.wikipedia.org/wiki/I-mode (Accessed: 3.Mai 2005)

[19] InfoSpace kauft in Deutschland zu, FinanzNachrichten,[online] 2004
http://www.finanznachrichten.de/nachrichten-2004-12/artikel-4206064.asp (Accessed: 3.Mai 2005)

[20] J. Hill, "Microsoft .NET Compact Framework", (Computerwoche), [online] 2002,
http://www.computerwoche.de/index.cfm?pageid=255&artid=36656&main_id=36656&category=160&currpage=2&type=detail&kw=
(Accessed: 3.Mai 2005)

[21] Java Native Interface, Wikipedia, [online] 2005,
http://de.wikipedia.org/wiki/JNI (Accessed: 3.Mai 2005)

[22] JavaOne 2004 - Java immer und überall, Sun, [online] 2005,
http://de.sun.com/homepage/feature/2004/javaone/ (Accessed: 3.Mai 2005)

[23] Kopierschutz für Handyklingeltöne und Java-Spiele, Heise, [online] 2002
http://www.heise.de/mobil/newsticker/meldung/29013 (Accessed: 3.Mai 2005)

[24] Khronos-Overview, Khronos, [online] 2005,
http://www.khronos.org/ (Accessed: 3.Mai 2005)

[25] Lexikon: Microsoft, ComputerBase, [online] 2005,
http://www.computerbase.de/lexikon/Microsoft#Positionierung (Accessed: 3.Mai 2005)

[26] M.Dupuis; B.Liebig ;P.Morandi, "Informatik in der Schweiz: Ausbildung, Beschäftigung, Markt (1981-2001)", (BBT), [online] 2003,
http://www.bbt.admin.ch/berufsbi/grund/feld/informatik/d/bericht_neu.pdf (Accessed: 3.Mai 2005)

[27] M. Juntao Yuan, "Challenges and opportunities in mobile games", (IBM), [online] 2004,
http://www-128.ibm.com/developerworks/wireless/library/wi-austingameconf.html (Accessed: 3.Mai 2005)

[28] M. Juntao Yuan, "Let the mobile games begin", (Javaworld), [online] 2002,
http://www.javaworld.com/javaworld/jw-02-2003/jw-0221-wireless.html (Accessed: 3.Mai 2005)

[29] Macrospace-Overview, Macrospace, [online] 2005,
http://www.macrospace.com/ (Accessed: 3.Mai 2005)

[30] MGain, "Mobile Entertainment Industry and Culture", (MGain), [online] 2003,
http://www.mgain.org/MGAIN-wp3-d311-revised-final.pdf (Accessed: 3.Mai 2005)

[31] Microsoft dominiert den Handheld-Markt, Winfuture, [online] 2004,
http://www.winfuture.de/news,17545.html (Accessed: 3.Mai 2005)

[32] MSDN Library, Microsoft, [online] 2005,
http://msdn.microsoft.com/library/default.asp (Accessed: 3.Mai 2005)

[33] Mobile2Market: How to Participate, Microsoft, [online] 2005,
http://msdn.microsoft.com/mobility/windowsmobile/partners/mobile2market/participatevendors.aspx (Accessed: 3.Mai 2005)

[34] Nokia, "Introduction to the Mobile Game Business", (Nokia), [online] 2003
http://nds2.forum.nokia.com/nnds/ForumDownloadServlet?id=2896&name=Introduction%5Fto%5FMobile%5FGames%5FBusiness%5Fv1%5F0%2Epdf (Accessed: 3.Mai 2005)

[35] Nokia, "Introduction to Mobile Game Development", (Nokia), [online] 2003
http://www.forum.nokia.com/html_reader/main/1,4997,2768,00.html?page_nbr=2#head2 (Accessed: 3.Mai 2005)

[36] Nokia, "Multiplayer Mobile Games: Business Challenge and Opportunities", (Nokia), [online] 2003
http://nds2.forum.nokia.com/nnds/ForumDownloadServlet?id=4796&name=business%5Fmultiplayer%5Fv%5F1%5F0%5Fen%2Epdf (Accessed: 3.Mai 2005)

[37] Nokia-Overview, Nokia, [online] 2003
http://www.nokia.ch/german/ (Accessed: 3.Mai 2005)

[38] Nokia, "Overview of Multiplayer Mobile Game Design", (Nokia), [online] 2004
http://nds2.forum.nokia.com/nnds/ForumDownloadServlet?id=4155&name=Multi%5Fplay%5FMobi%5Fv1%5F1%5Fen%2Epdf (Accessed: 3.Mai 2005)

[39] NTT Docomo, Wikipedia, [online] 2005,
http://de.wikipedia.org/wiki/Docomo (Accessed: 3.Mai 2005)

[40] Overview, Symbian, [online] 2005,
http://www.symbian.com/about/about.html (Accessed: 3.Mai 2005)

[41] P. Yao, "Microsoft .NET Compact Framework for Windows CE .NET", (Microsoft), [online] 2002,
http://msdn.microsoft.com/library/default.asp?url=/library/en-us/dncenet/html/wincecompactfx.asp (Accessed: 3.Mai 2005)

[42] PalmOne-Overview, PalmOne, [online] 2005
http://software.palmone.com/ch/de/PlatformSoftware.jsp?siteId=1017&platformId=1 (Accessed: 3.Mai 2005)

[43] PalmSource-Overview, PalmSource, [online] 2005,
http://www.palmsource.com/ (Accessed: 3.Mai 2005)

[44] Pressemappe Detecon Consulting, Presseportal, [online] 2004,
http://www.presseportal.de/story.htx?nr=552849&firmaid=42555 (Accessed: 3.Mai 2005)

[45] PC- & Konsolengames: Halflife2, Pro7, [online] 2005,
http://www.prosieben.de/games_handy/pc_konsolengames/specials/halflife2/ (Accessed: 3.Mai 2005)

[46] Products, Samsung by Samtel AG, [online] 2005,
http://www.samtel.ch/main/d/produkte/index.html (Accessed: 3. Mai 2005)

[47] QUALCOMM-Overview, QUALCOMM, [online] 2005,
http://www.qualcomm.com/about/index.html (Accessed: 3.Mai 2005)

[48] S. Kangas, "Mobile Entertainment Industry and Culture", (MGain), [online] 2003,
http://www.mgain.org/mgain-wp6-d6214.pdf (Accessed: 3.Mai 2005)

[49] SDN Mobility Program Go-To-Markte Services, Sun, [online] 2005,
http://developers.sun.com/techtopics/mobility/developers/mobility/gotomarket/javaverification/ (Accessed: 3.Mai 2005)

[50] Siemens-Overview, Siemens, [online] 2005,
http://www.siemens.ch/index.jsp (Accessed: 3.Mai 2005)

[51] Sony Playstation, Sony, [online] 2005,
http://de.playstation.com/news/newsStory.jhtml?storyId=106718_de_DE_NEWS&linktype=NT3 (Accessed: 3.Mai 2005)

[52] SonyEricsson-Overview, SonyEricsson, [online] 2005,
http://www.sonyericsson.com/spg.jsp?cc=ch&lc=de&ver=4000&template=ph1&zone=ph (Accessed: 3.Mai 2005)

[53] Sun, "Java 2 Platform, Micro Edition", (Sun), [online] 2002,
http://java.sun.com/j2me/j2me-ds.pdf (Accessed: 3.Mai 2005)

[54] Sun, "J2ME Building Blocks for Mobile Devices", (Sun), [online] 2000,
http://java.sun.com/products/cldc/wp/KVMwp.pdf (Accessed: 3.Mai 2005)

[55] Sun, "Mobile Information Device Profile", (Sun), [online] 2003,
http://java.sun.com/products/midp/midp-ds.pdf (Accessed: 3.Mai 2005)

[56] Sun, "The network connection", (Sun), [online] 2002
http://developers.sun.com/techtopics/mobility/midp/chapters/j2mewhite/chap13.pdf (Accessed: 3.Mai 2005)

[57] Symbian baut Marktposition aus, ZDNet, [online] 2004,
http://www.zdnet.de/news/software/0,39023144,39124643,00.htm (Accessed: 3.Mai 2005)

[58] Symbian-Overview, Symbian, [online] 2005,
http://www.symbian.com (Accessed: 3.Mai 2005)

[59] T. Obermaier, "J2ME Game Development", (Games-Net), [online] 2002,
http://www.games-net.de/files/j2me.pdf (Accessed: 3.Mai 2005)

[60] Ubisoft-Overview, Ubisoft, [online] 2005
http://www.ubisoftgroup.com (Accessed: 3.Mai 2005)

[61] W. Uyttersprot, "Let's go compact: the .NET Compact Framework", (U2U), [online] 2003,
http://www.u2u.net/ArticlePage.aspx?ART=msdnetcompact-e (Accessed: 3.Mai 2005)

[62] Why Develop on the Windows Mobile Platform, Microsoft, [online] 2005,
http://msdn.microsoft.com/mobility/windowsmobile/why/whydevelop/default.aspx (Accessed: 3.Mai 2005)

[63] Windows Mobile, Microsoft, [online] 2005,
http://www.microsoft.com/windowsmobile/devices/default.mspx (Accessed: 3.Mai 2005)

[64] World of Warcraft Aktuelle Verkaufszahlen, Game7, [online] 2005,
http://www.game7.de/223-world-of-warcraft/n052si2110.php (Accessed: 3.Mai 2005)

# Chapter 8

# Security in WLAN

*Samuel Förstler und Michael Müller*

*Wireless LANs are not sufficiently protected, even if network administrators use the built-in security protocol WEP (Wired Equivalent Privacy). In 2002, a seven-month investigation conducted in London found that ninety-four percent of all wireless LANS in use were inadequately protected from attacks. It was found that wireless networks are poorly secured, if they are secured at all. Drive-by hacking becomes more and more popular: hackers drive through office districts in a car and try to penetrate company WLANs from the street where their signals can still be received. In this paper we intend to discuss on the one side several security problems which arise in wireless networks. On the other side we like to present a number of different countermeasures according to this problem.*

## Contents

# 8.1 Introduction

## 8.1.1 WLAN: A short introduction

Wireless LAN (WLAN), based on the defined standard IEEE 802.11 from 1997 of the Institute of Electrical and Electronics Engineers (IEEE), offers the possibility of developing wireless local networks or to extend existing wire-bound networks at small expenditure.

The basic configuration of wireless LAN is not much different from that of wired LAN. The biggest difference is that wireless LAN uses radio waves, while conventional LAN uses wires and cables. By adopting the use of radio waves, the range of communications is much farther than with infrared rays, with much less interference. Because of a huge frequency gap between mobile or wireless phones and wireless LAN, which uses 2.4GHz, there is no risk of crossing transmissions with other communication devices. There is also no need to apply for a license.

Due to the simple installation, WLAN is used also for networks which can be installed only temporarly (e.g. on exhibitions). Further, it brings the possibility to offer network entrances, so called Hot Spots, at public places such as airports or train stations in order to make it possible for mobile users to establish connections to the Internet or to the Home Office.

Since 2001 however, safety gaps are well-known in the standard, which can lead to serious safety problems.

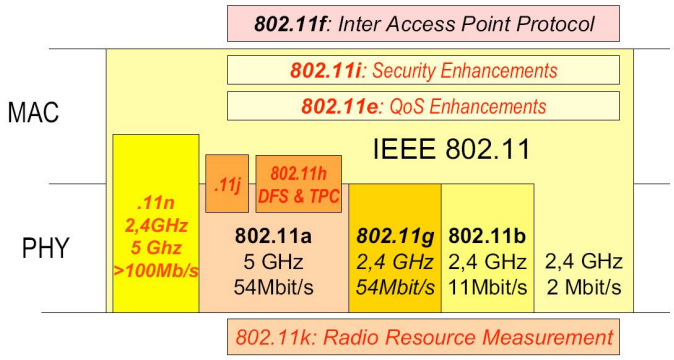


Figure 8.1: Overview of the IEEE802.11 standard

## 8.1.2 Physical Network vs. WLAN: Different aspects of security

These problems arise from the main difference of WLANs to the usual data communication in local area networks, the fact that instead of cables, the transference medium of communication is the air (see fig. 8.2). Due to the propagation characteristic of electromagnetic waves, hearing or sending on the physical level is possible without a physical break into a building (Parking Lot Attack). That is a substantial difference to wire-bound transmission protocols which are used in physical networks. Using this kind of network,

only the data traffic which leaves the internal system can be observed from outside of a building.

To make it more difficult, it has to be added that WLAN systems should make it possible to install a comfortable Ad-hoc-Networking. Thereby, the identification, authentication and registration (authorizing) of the stations must run off as automatically as possible. Usually, the WLAN systems available at the market contain safety mechanisms, but they have partially substantial security. Attacks are very possible.

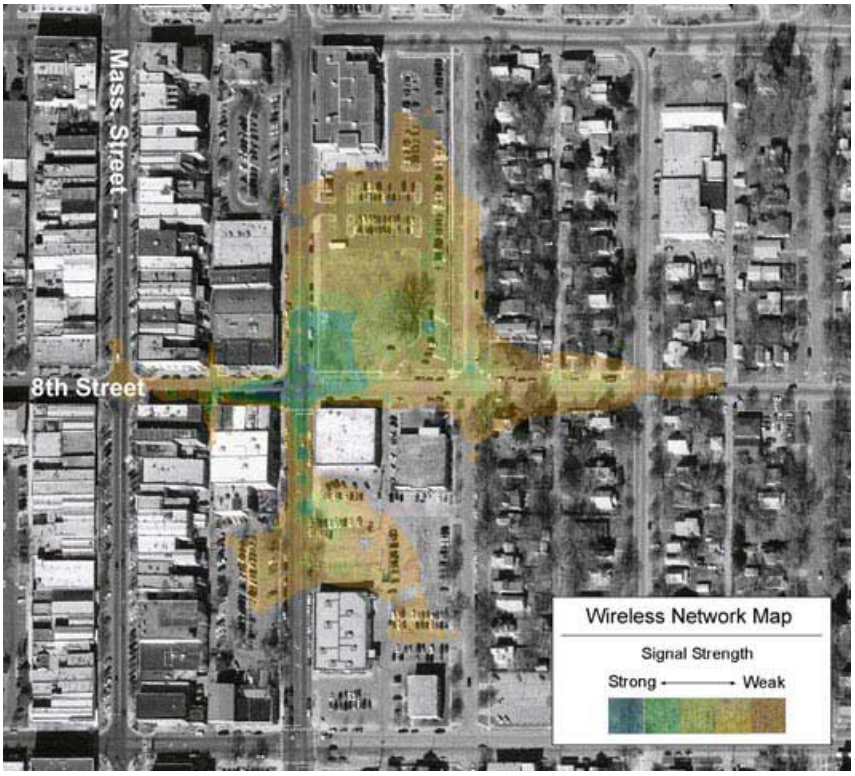


Figure 8.2: Area which is covered by radiosignals from a WLAN.

### 8.1.3 Architecture

IEEE 802.11 describes the concept of the Independent Basic Service Set (BSS) which is in effect the Wireless LAN subset consisting of the clients and perhaps an Access Point.

In principle, WLAN can be operated in two different types of architecture. First, the Ad-hoc-mode: Two or more mobile terminals (Clients) which are equipped with a WLAN map, communicate directly with one anotherd (see fig. 8.3).

In most cases, WLAN is operated in infrastructure mode, i.e. the communication of the Clients is realized by a central radio bridge, the so-called Access Point. If cable-bound LAN segments are part of the system, this access is also realized by the Access Point (see fig. 8.4).

Figure 8.3: Ad-Hoc-Installation


Figure 8.4: Infrastructure Modes

## 8.2 Security Mechansims

### 8.2.1 Security - A short definiton

As with any network communication system, the basic security objectives for securing a wireless LAN involve confidentiality, integrity, and access control.

#### Confidentiality

Prevents eavesdropping and ensures only the intended audience understands the transmitted data. The standard solution for this issue is encryption.

#### Data Integrity

Prevents attackers from transparently modifying data and ensures the data comes from the source it purports to come from. Solutions for this problem are often realized with checksum functions.

**Access Control**

Prevents unauthorized users from communicating through the wireless access points. Access control also ensures that legitimate clients associate with trusted, rather than 'rogue' access points. To provide these services, WLAN security uses Service Set Identities (SSIDs), symmetric cryptography and checksum functions which are discussed below.

### 8.2.2 Security Mechanisms in WLAN

**Basic Mechanisms**

**SSID** The first of these technologies is called the Service Set Identifiers (SSID). The SSID is a basic naming handle and is a network name for a set of devices in a Wireless subsystem. The SSID logically segments the Wireless LAN. The SSID differentiates one WLAN from another, so all access points and all devices attempting to connect to a specific WLAN must use the same SSID.

**ACL** The 802.11 MAC layer provides functionality to allow reliable data delivery for the upper layers over the wireless PHY media. The data delivery itself is based on an asynchronous, best-effort, connectionless delivery of MAC layer data. There is no guarantee that the frames will be delivered successfully.

The 802.11 MAC provides a controlled access method to the shared wireless media called Carrier-Sense Multiple Access with Collision Avoidance (CSMA/CA). CSMA/CA is similar to the collision detection access method deployed by 802.3 Ethernet LANs.

The third function of the 802.11 MAC is to protect the data being delivered by providing security and privacy services. Security is provided by the authentication services and by Wireless Equivalent Privacy (WEP), which is an encryption service for data delivered on the WLAN.

**WEP**

IEEE802.11 defines an optional WEP mechanism to provide confidentiality and integrity of traffic in the WLAN. WEP is used at the two lowest layers of the Open Systems Interconnect (OSI) reference model, data link and physical layers; thereby, it does not offer end-to-end security.

The key elements of WEP are:

- Secret WEP-Key
- Initialization Vector (IV)
- RC4 PRNG algorithm

- Integrity check vector (ICV)

WEP depends on a secret Key shared between the communicating parties (mobile station and AP) to protect the payload of a transmitted frame in each direction. For this operation, also a Initialization Vector (IV) is used. Moreover, the RC4 PRNG algorithm used by WEP includes an integrity check vector (ICV) to check the integrity of each packet. This process is summarized below (see fig. 8.5).

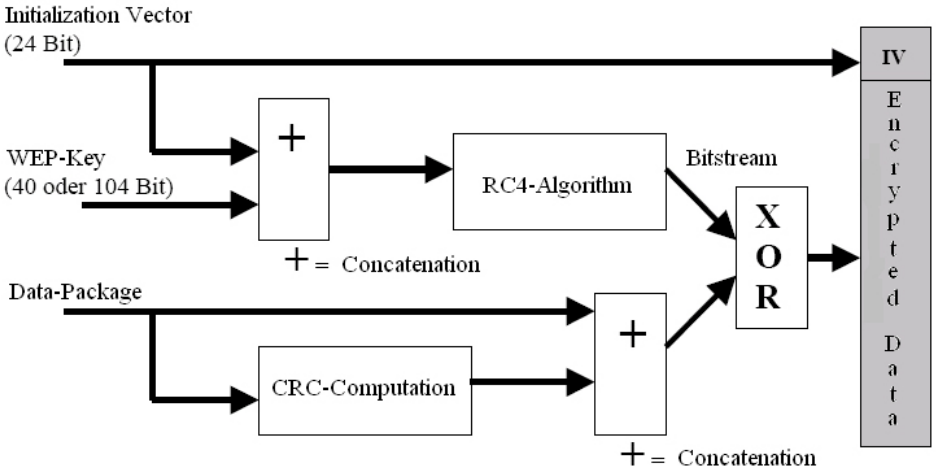


Figure 8.5: Encryption-scheme

**Authentication in WEP** Authentication of the client can happen in either of two ways, an Open key method, or a Shared key method. Whichever method is used must be the same for the whole subsystem i.e. the devices and the AP.

**Open Key** tThis is the default mechanism and allows the whole authentication process to occur openly in clear text so that any client can associate with the AP, even if they supply a wrong key.

**Shared Key** The AP sends a challenge text packet to the client. The client must respond with the text encrypted with the correct WEP key. Some systems require a MAC address to be used with the response so that a client's MAC address must match one that has been previously entered into the APs association table (see fig. 8.6).


Figure 8.6: Authentication Scheme

**Confidentiality and Integrity in WEP** First, WEP computes the ICV by performing a 32-bit cyclical redundancy check (CRC-32) of the frame and appends the vector to the original frame, resulting in the plaintext. Then, the message plus ICV is encrypted via the RC4 PRNG algorithm using a long sequence key stream, a long sequence of pseudorandom bits. This key stream is a function of the 40-bit secret key (which is shared between all authorized stations in the WLAN) and a 24-bit initialization vector (IV). Consequently, an exclusive-or (XOR) operation is made between the plaintext and the key stream to produce the ciphertext. Finally, it is the ciphertext that is sent over the radio link. Theoretically, the ciphertext provides data integrity because of the ICV and confidentiality due to encryption.

The receiver, inasmuch as RC4 PRNG algorithm is symmetric, performs the same procedure described above, but in reverse, to retrieve the original message frame. Specifically, the ciphertext is decrypted using a duplicated key stream to recover the plaintext. The recipient then validates the checksum on this plaintext by computing the ICV and comparing it to the last 32 bits of the plaintext, thus ensuring that only frames with a valid checksum will be accepted by the receiver.

WEP can be implemented with the classic 40-bit key and 24-bit IV or a vendor-dependent (hence proprietary) extended version that affords a larger key. The shorter key length can be relatively easy to compromise via brute-force attack, even with modest computing resources; however, a larger key such as the 128-bit keys would be render brute-force attacks impossible, even for sophisticated computing systems. Nevertheless, alternative attacks are possible that do not require a brute-force strategy, thereby diminishing the strength of key length.

### WPA

**WPA** (Wi-Fi Protected Access) is a system to secure wireless (Wi-Fi) networks, created to patch the security of the previous system, **WEP** (Wired Equivalent Privacy); researchers have found a number of weaknesses in WEP. As a successor, WPA implements the majority of the IEEE 802.11i standard [4], and was intended as an intermediate measure to take the place of WEP while 802.11i was prepared. WPA was created by The Wi-Fi Alliance, an industry trade group, which owns the trademark to the Wi-Fi name and certifies devices that carry that name. Before we start with our detailed look on WPA we first give a short overview over the different operational modes (see fig. 8.7).

| *Mode* | *WPA* | *WPA2* |
|---|---|---|
| Enterprise Mode<br>(Business and Government) | Authentication: IEEE 802.1X/EAP<br>Encryption: TKIP/MIC | Authentication: IEEE 802.1X/EAP<br>Encryption: AES-COMP |
| Personal Mode<br>(SOHO/personal) | Authentication: PSK<br>Encryption: TKIP/MIC | Authentication: PSK<br>Encryption: AES-COMP |

Figure 8.7: WPA Modes

For now, we ignore the **WPA2** implementation mentioned in the third column and focus on **WPA**. We will discuss **WPA2** later in this paper. As we can see, there exist two

operational mode [4] The first, and also the more important one, is the **Enterprise-Mode**. When we look on our security framework with its three pillars of security aspects, namely authentication, encryption, and integrity, we also find these aspects in the WPA implementation. Here a short overview:

**WPA-Concept**

1. **Authentication:**

   - Pre-shared Key for personal mode (less secure) or
   - IEEE802.1X(AAA-Server)/ EAP in the enterprise mode. Today, there exists a lot of providers for AAA-Servers. We have **Kerberos**, **RADIUS**, **free RADIUS** and a lot more. These servers are embedded in a security concept that uses a special protocol for user and server side identification. But we will describe this Protocol (EAP) further down. AAA (or triple-A) in this context is a shortcut and stands for **A**uthentication, **A**uthorization and **A**ccounting. As we will see, these servers are connected with a authentication database for checking and verifying the participant's id.

2. **Privacy:**

   - TKIP (pronounced tee-KIP) is the temporal key integrity **protocol** and is a wrapping for the WEP algorithm to hide several weak points in the RC4 implementation. This protocol uses a 128-bit key length for its per-packet keying.

3. **Data integrity:**

   - WPA uses for its integrity check a separate algorithm, the MIC (message integrity check). Each side (server/client) uses a key with a 64-bit length for integrity.

For the authentication aspect we have a IEEE 802.1X respective an EAP implementation to grant authentication and authorisation. **EAP** stands for "'Extensible Authentication Protocol"' and provides any need for mutualy identification. As already mentioned, it is a protocol from the IEEE 802-standard family with a port-based access control. So that we can use EAP, WPA is designed for use with an 802.1X authentication server, which distributes different keys to each user; however, it can also be used in a less secure pre-shared key (PSK) mode, the second operational mode in our table showed above. This mode is also known as the personal mode or SO/HO (small office / home office) mode.

For data encryption, our second pillar in the security framework, WPA is using the RC4 stream cipher, with a 128-bit key and a 48-bit initialization vector (IV). As we see, key length an IV length has been increased for higher privacy. One major improvement over WEP is given by the Temporal Key Integrity Protocol (TKIP), which dynamically changes keys as the system is used. When combined with the much larger IV, this defeats the well-known key recovery attacks on WEP. But as already mentioned, the encryption still

bases on the RC4 algorithm with its well known weake points and the TKIP just enwrap this algorithm. For more information about TKIP, read the subsection TKIP (see 8.2.2).

In addition to authentication and encryption, WPA also provides vastly improved payload integrity, our third pillar of the security framework. The cyclic redundancy check (CRC) used in WEP is inherently insecure; it is possible to alter the payload and update the message CRC without knowing the WEP key. A more secure message authentication code (here termed a Message Integrity Check (MIC)) called "'Michael"' is used in WPA. Further, the MIC used in WPA includes a frame counter, which prevents replay attacks being executed.

WPA was specifically extracted as an intermediate step towards improved 802.11 security for two reasons: first, 802.11i's work lasted far longer than originally anticipated, spanning four years, during an period of ever-increasing worries about wireless security; second, it encompasses as a subset of 802.11i only elements that were backwards compatible with WEP for even the earliest 802.11b adapters. WPA firmware upgrades have been provided for the vast majority of wireless NICs ever shipped; 802.11 gateways sold before 2003 generally needed to be replaced. By increasing the size of the keys, the number of keys in use, and adding a secure message verification system, WPA makes breaking into a Wireless LAN far more difficult. The Michael algorithm was the strongest that WPA designers could come up with that would still work with most older network cards; however it is subject to attack. To limit this risk, WPA networks shut down for 30 seconds whenever an attempted attack is detected.

WPA2 is the certified form of IEEE 802.11i tested by the Wi-Fi Alliance. WPA2 implements the mandatory elements of 802.11i. In particular, the Michael algorithm is replaced by a message authentication code that is considered fully secure. The Wi-Fi Alliance calls the pre-shared key version WPA-Personal or WPA2-Personal and the 802.1X authentication version WPA-Enterprise or WPA2-Enterprise. The group uses these terms to allow purchasers to know at a glance what the highest level and type of protection is available with a device.

### TKIP

The Temporal Key Integrity Protocol TKIP is part of the IEEE 802.11i encryption standard for wireless LANs [4], [1]. TKIP (pronounced tee-KIP) is the next generation of WEP, the Wired Equivalency Protocol (or Wired Encryption Protocol), which is used to secure wireless LANs based on the 802.1X-Standard and also by WPA for authentification. This advanced encryption was developed by the IEEE 802.11 Task Group i (TGi). TKIP intends to quickly fill the gaping hole left by Wired Equivalent Privacy. This means TKIP provides per-packet key mixing, a message integrity check and a re-keying mechanism, thus fixing the flaws of WEP. Therefor TKIP always changes its dynamic keys between AP and client after a set of packets has been sent (after about 10'000 packets).

An other requirement for this enhancement of the insecure WEP-encryption was the downward compatibility with popular legacy equipment. Thus the intend during development was that it just can be deployed in firmware updates instead of new chipsets for protecting the consumer investment.

### EAP

EAP 'Extensible Authentication Protocol' is a universal authentication mechanism, frequently used in wireless networks and Point-to-Point connections. When invoked by an 802.1x enabled NAS (Network Access Server) device such as an 802.11 a/b/g Wireless Access Point, it can provide a secure authentication mechanism and negotiate a secure PMK (Pair-wise Master Key) between the client and NAS for strong encryption. As we see in the table, there exist a lot of different implementations. In this paper we will discuss the highlighted protocol called EAP-TLS (see fig. 8.8). It works on a secure transport layer (**T**ransport **L**ayer **S**ecurity).

| Protocol | Client/Server-Authentication | dynamic WEP Key management |
|---|---|---|
| EAP-MD5 | Nein | Nein |
| EAP-TLS | Ja | Ja |
| EAP-TTLS | Ja | Ja |
| EAP-PEAP | Ja | Ja |

Figure 8.8: EAP Protocols

### EAP-TLS

As already mentioned, EAP-TLS is a transport protocol for authentication informations between a client and in our case an access point. Authentication takes place over a AAA-Server with a database where certificates and user-id are saved. EAP-TLS in our case is very interesting because we need a secure communication during protocol an key negotiation over the air. Fortunately TLS is implemented in the OSI-architecuter directly above the TCP/IP-Layer. So any Session which uses TCP can benefit from TLS. TLS is similar to SSL (Secure Socket Layer) but they aren't substitutable. For an overview, we look at the next fig. 8.9 and discuss step-by-step the sequence.

EAP-TLS Sequence:

1. **EAP-Start:** After the EAP-start the authenticator sends a message back to the supplicant.

2. **EAP-Request/Identity:** The authenticator wants the UserID, Password and Certificate from the supplicant

3. **EAP-Response/Identity:** The supplicant answers with the wanted information

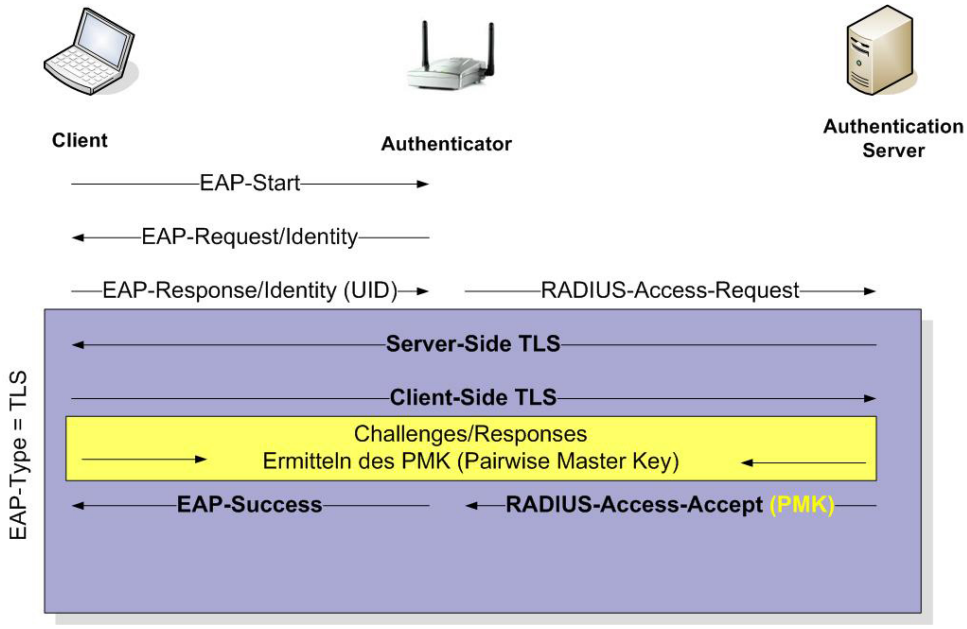


Figure 8.9: EAP-TLS Sequence

4. **RADIUS-Access-Request:** The authenticator forwards the supplicant's message with a Radius Access Request to the Authentication-Server.

   **TLS-Authentication begins:**

5. **Server-Certificate:** The server sends its certificate to the client and asks for the supplicant's certificate.

6. **Client-Certificate:** After checking the server's cert the client sends its own cert to the server.

7. **Negotiating:** The server checks also the supplicants cert and dertermines the session specific parameters.

8. **Session Key:** Parallel, server and client determine independently the session key.

9. **Radius-Access-Success:** This session key will sent with a Radius Access Success message to the authenticator and checks wether supplicant and authenticator work with the correct keys and session parameters.

10. **EAP-Success:** The authenticator sends a EAP-Success message to the client and data communication can start.

## 8.3 Criticism on the most popular Security Mechanisms

### 8.3.1 Basic Mechanisms

#### SSID

The service set identifier was originally created as a minimal security measure. It is not currently regarded as a security measure since it is analogous to a user name without any password. Moreover, this 'user name' or SSID is beaconed (transmitted) by the access point for anyone to grab onto and join your network. That means that an SSID can be sniffed in plain text from a packet and does therefore not supply any security to the network.

### 8.3.2 ACL

One Disadvantage is, that the own infrastructure has to be known. This results in a very static network, because each computer that will connect to the net must have a MAC-address enlisted by the network administrator.

Each network card in a computer has its own MAC-address given by the hardware producer. With simple software tools from the internet, the unique MAC-address can be changed in the designated hex-address. But how do you get a valid MAC-address accepted by the invaded network? The mentioned address is saved separately in a ROM-part on the network-card and can't be hacked from outside the net. But when data is sent in packets over access points the MAC-address is not encrypted and can easily be catched.

### 8.3.3 WEP

As mentioned above, WEP has also some disadvantages. The most important of them are discussed in this section.

- Optional and disabled security features

Security features, albeit poor in some cases, are not enabled when shipped, are user don't enable when installed. But bad security is generally better than no security at all.

- Short and static IVs

24-bit IVs cause the generated key stream to repeat. Repetition allows easy decryption of data for a moderately sophisticated adversary.

- Short cryptographic keys

40-bit keys are inadequate for any system. It is generally accepeted that key sizes should be greater than 80 bits in length. The longer the key, the less likely a comprise from a brute-force attack.

- Shared cryptographic keys

Keys that are shared can compromise a system. As the number of people sharing the keys grows, the security risks also grow. A fundamental tenant of cryptography is that the security of a system is largely dependent on the secrecy of the keys.

- No frequent and automatic update for cryptographic keys

Cryptographic keys should be changed often to prevent brute-force attacks.

- Weak RC4

The combination of revealing 24 key bits in the IV and a weakness in the initial few bytes of the RC4 key stream leads to an efficient attack that recovers the key. Most other applications of RC4 do not expose the weakness of RC4 because they do not reveal key bits and do not restart the key schedule for every packet. This attack is avaitable to moderately sophisticated adversaries.

- Poor packet integrity

CRC32 and other linear block codes are inadequate for providing cryptographic integrity. Message modification is possible. Linear codes are inadequate for the protection is requird to prevent deliberate attacks. Use of nonoryptographic protocols often facilitates attacks against the cryptography.

- Weak and incomplete authentication

Only the device is authenticated. A device that is stolen can access the network. Further, Authentication is not enabled; only simple SSID indentification occurs. Identity-based systems are highly vulnerable particularly in a wireless system because signals can be more easily intercepted. Device authentication is simple shared-key challenge-response. One-way challenge-response authentication is subject to 'man-in-the-middle' attacks. Mutual authentication is required to provide verfication that users and the network are legitimate. The client does not authenticate the AP. The client needs to authenticate the AP to ensure that it is legitimate and prevent the introduction of rogue APs.

### 8.3.4 WPA

**The TKIP** still bases on the weak RC4-algorithm which is not very safe. As we have seen, safety has been encreased by extending the key's and inital vector's length and as major improvement the protocol does not use a static key.

**Trivial Passwords** make better protection increased by key length and per-packet key management obsolete.

**The ad-hoc mode** is still unprotected

Because of all this disadvantages, there was a need of some improvements concerning the security mechanisms in WLAN. New concepts were (and are) developped. One of the most important organization behind this topic is the WIFI-Alliance.

### 8.3.5 WiFi-Alliance

The WIFI-Alliance is a membership organization founded in 1999 devoted to certifying the 802.11 wireless Ethernet devices for interoperability. The Alliance is addressing the need for an immediate, software-upgradeable security solution. Realizing the importance of enhanced Wi-Fi security, the Alliance has led an effort to bring strongly improved, interoperable Wi-Fi security to market early this year. The result of that effort is Wi-Fi Protected Access.

Wi-Fi Protected Access is a specification of standards-based, interoperable security enhancements that strongly increase the level of encryption and authentication for existing and future wireless LAN systems. Wi-Fi Protected Access is derived from the upcoming IEEE 802.11i standard and will be forward compatible with it.

Wi-Fi Protected Access addresses the vulnerabilities of WEP encryption and adds user authentication. Thus, Wi-Fi Protected Access will provide wireless LAN users with a high level of assurance that their data will remain protected and that only authorized network users can access the network. Significantly, it is designed as a software upgrade to Wi-Fi CERTIFIED devices, requiring no additional hardware. Wi-Fi Protected Access includes 802.1X and TKIP technology. Cryptographers working with the Wi-Fi Alliance have reviewed Wi-Fi Protected Access and endorsed the fact that it solves all of WEP's known vulnerabilities. Wi-Fi Protected Access support is available from vendors of WLAN equipment since early 2003.

### 8.3.6 Conclusion

As we have seen so far, all showed solutions have their advantages but also their disadvantages. But any of these solutions doesn't offer full protection from attack by hackers. So IEEE TaskGroup has started to develope a new security concept. But as we know, those concepts need time to get realized. Therefore the WI-FI Alliance offerd an interim

solution the WPA. They intended to realize a simple answer to the given problems which can work together with the existing consumer hardware. So any changes in the existing WEP concept had to be done on the software level but that still wasn't secure enough. What's coming up then? This we will discuss in the chapter 8.5. But first we will see, how we can protect ourselves with our given hardware and how we can increase our privacy with little effort.

## 8.4 Measures to Encrease Security in WLAN

We heard a lot about gaps and problems in the security setup for wireless LAN so far. But what can we do now, with the given solutions to improve our privacy and safety during the network session? The countermeasures can be devided in three parts [1]that should be kept in mind and can increase your safety enormously. The next three subchapters

- Configuration and Maintenance
- Additional Technical Measures
- Organisational Measures

will explain in detail what exactly can be done.

### 8.4.1 Configuration and Maintenance

Despite of all problems in the above discussed countermeasures, we should activate these security settings to prevent attacks by disposable tools by the internet. This means in detail:

1. **Basic Protection Mechanisms:**
   - change the AP's and client's standard password
   - change the standard SSID of the AP
   - disable SSID broadcast
   - enable MAC-address filtering with the ACL
   - enable the WEP encryption (128 bit if supported)
   - use an open WEP key because 'shared key' has additional security holes
2. **Change the WEP key periodically.**
   Therefor you should follow the common rules to create a password
3. **Optimate the way the trasmitter is sending.**
   You should keep the transmitting range as small as possible, so that only those areas has access to the network, which are allowed

4. **Deactivate the DHCP server on your access point.**
   Your access point's DHCP server should be deactivated if possible and you should use static IP adresses in a short range

5. **Upgrade the firmware immediately on the hardware after its release by the hardware manufacturer.**
   With this update your AP will get enhanced proprietary security options that can improve security in case of compatibility.

6. **Turn off the transmitter if it is not in use.**
   This applies also for clients because of the ad-hoc mode.

7. **Maintain the access point only over secured channels.**
   Otherwise sensitive settings can be catched by others

### 8.4.2 Additional Technical Measures

In addition to the 802.1X standard the following technical measures are essential to increase privacy.

1. **The use of additional security solutions:** Each three parts of security (authentication, privacy and integrity) should be combined together in the best way. But there ever will be a trade-off between security and flexibility. The more secure a system is the more inflexible it gets. If you want to use the more secure authentication principle of EAP for example (instead the pre-shared key) you need more hardware and a more complex user account administration and this makes it more difficult to setup a spontaneous session. An other solution could be a VPN. The VPN has to be installed behind an AP and each time a client establishes a connection to the network between this client and the VPN-gateway it occurs a tunneling [1].

2. **The separation of the physical LAN:** The physical network (especially important for government and enterprises) should be strictly separated from APs by Firewalls and intrusion detection systems (IDS).

3. **Protection for clients:** Especially mobile clients as notebooks oder PDAs should implement local protection measures as access control, user authentication (fingerprints), antivirus protection, personal firewalls, restrictive access to hardware ressources (e.g. harddrive access) and local data encryption.

### 8.4.3 Organisational Measures

These three non-technical measures [1] help to increase quality of the already mentioned technical arrangements in section 8.4.1 and 8.4.2.

1. Release security guidelines

2. Observe the compliance with the existing security guidelines
3. Protect your personal data when using unknown hotspots

## 8.5 Perspectives

In this chapter, we will have a look at the security standard IEEE 802.11i and how it is implemented. As we have seen, a few concepts of this new standard have already been impelmented in the WPA architecture, because there was a need for more security during the time between the WEP solution and the release of the new 802.11i standard [4]. In September 2004 the IEEE 802.11i was finally released and the WIFI-Alliance has started to certify the new hardware supporting WPA2. In the next chapter we will see, why they have to certifiy new hardware and why we can't use the hardware based on 802.11g and older standards.

### 8.5.1 A New Standard WPA2

Both standards (WPA and WPA2) support the enterprise and personal mode. For a better overview please look up at (see fig. 8.7). As we can see nothing changed on the level for authentication. In the Enterpires mode they suggest to use a port-based authentication protocoll also known as EAP (see Section 8.2.2). The use of a pre-shared key wouldn't be very reasonable. If we had to share our key with many employee in our firm, the key would be an open secret and the security concept could not grant the correct authentication. On the other hand, we have to pay for this enhanced security with less flexibility, because we have to manage authorisations, certificates and levels of access permission.

One of the major update is the protocol for encryption. The IEEE TG determined that their approach to realize security encryption on the physical layer was wrong. They had to implement the new solution an a higher level in collaboration with the hardware, because encryption by software, wasn't secure at all. Each time they patched weak holes in the encryption algorithm, after a short time, the hacker scene answered with new tools and tipps to attack other hidden weak security holes. Therefore the implementation had to occur in hardware. And the underlaying concept is the **AES-CCM** (**A**dvanced **E**ncryption **S**tandard) protocol. This protocol we will discuss in section 8.5.2.

Other improvements which have been implemented:

1. They now offer a security support for the ad-hoc mode (see fig. 8.3).
2. Secure Fast Handoff
3. pre-Authentication: this works hand in hand with the secure fast handoff. A handoff occurs when a mobile node exits the AP's transmission range and changes to another AP (transmission ranges may overlap). A special protocol has to establich a new packet routing to the actual AP and has to manage the problem of packet loss. With

pre-Authentication a mobile node has the ability to scan the transmission area and locate other APs. If there are other AP in range, the mobile node is able to carry out a pre-Authentication in case the mobile node will change the AP at a later date.

4. And at least a secure logoff from the net. This is important for enterprise networks with an EAP environment.

### 8.5.2 New Security Protocols - The Advanced Encryption Standard

The newest protocol for encryption is AES the Advanced Encryption Standard. This new protocol won in a contest by the NIST the National Institute of Standards an Technology. The requirements for the wanted security protocol were:

- AES has to be a block cypher algorithm with a symmetric key (same key for en- and decryption) .
- AES has to use a block size of 128 bit and a key size of 128, 192 or 256 bit.
- AES should have a little need in hardware resources
- And of course the most important: it should be robust

The Advanced Encryption Standard (AES) was intended to replace Triple DES, itself an interim fix for the aging Data Encryption Standard (DES). The primary motivation for a new standard was the fact that DES has a relatively small 56-bit key which was becoming vulnerable to brute force attacks. In addition the DES was designed primarily for hardware and is relatively slow when implemented in software. While Triple-DES avoids the problem of a small key size, it is very slow in software, and also unsuitable for limited-resource platforms.

The AES uses the CCM-Protocol. This protocol satisfy two requirements in our security framework. First, we have the Counter-Mode Protocol that is apropriate for encryption, and uses a a key with a size of 128 bit (determined by the WiFi-Alliance). Second, we have the CBC-MAC (Cipher Block Chain Message Authentication Code) for data integrity. The interestin thing in this new protocol is we have only one key and with this key we generate the others for encryption and integrity check.

### 8.5.3 Outlook

#### New Perspectives on WLAN

The slowdown in the implementation of 3G services is a leading example of the new reality of market saturation and increasingly sophisticated consumers. But, while operators struggle to bring UMTS services to the market in service configurations subscribers want and at prices they are prepared to pay, the explosive growth of WLAN offers new ways forward which, in many ways, support the delayed 3G revolution.

**The Facts**

If the mobile phone has had a massive influence on the way we live and organise our lives, mobile computing has had an equally massive effect on the way we work and where we work. End-users demand high speed services and 'The office is wherever I am' is the new reality.

This is supported by hard facts - a recent survey in an international business magazine shows that 50% of business travellers work in their hotel, 40% work in 'foreign' work places, usually the offices of clients or suppliers, and 30% work during travel.

Thus, mobile operators, with their very large subscriber bases, especially business subscribers, are very well placed to benefit from the growth opportunities of WLAN.

What is equally important is that the business segment will define both usage patterns and expectations for future data services.

**Challenges**

The commercial challenges involved in WLAN roaming include a current absence of industry standards, no suitable standard roaming agreement format, different settlement procedures and incompatible billing systems.

We have seen, there are problems with authentication methods, a lack of consistency in the user experience, beginning with the end user log-on, diverse information formats and big challenges in developing a 'one bill' system.

But industry has recognised the obstacles to WLAN roaming and is adapting knowledge to provide new industry standards for WLAN - uniting the different communities and not only making roaming technically possible but also making it commercially viable.

# Bibliography

[1] Bundesamt für Sicherheit in der Informationstechnik, Sicherheit im Funk-Lan, Juli 2002

[2] Deploying Wi-Fi Protected Access (WPA) and WPA2 in the Enterprise, Wi-Fi Alliance, März 2005

[3] IEEE 802 LAN/MAN Standards, IEEE Group, 2004

[4] Securing Wi-Fi Wireless Networks with Today's Technologies, (C) Wi-Fi Alliance, 2003

[5] Security for 802.11 Wireless Networks, Cherita Corbett, Dept. of Electrical and Computer Engineering / Georgia Institute of Technology

# Kapitel 9

# Security in GSM and GPRS Networks

*Gregor Berther und Beat Affolter*

*Die Mobiltelefonie ist ein immer integralerer Bestandteil unseres Lebens. Täglich führen wir Tausende von Telefongesprächen mittels GSM (Global System for Mobile Communications) und verschicken Millionen von Kurzmitteilungen. Auch die Datenübertragung über WAP (Wireless Application Protocol) oder GPRS (General Packet Radio Service) wird immer wichtiger. Doch bei aller Alltäglichkeit dieser Dienste fehlt den meisten Benutzern die Sensibilität für die Sicherheit ihrer Gespräche oder Daten. Mit dieser Seminararbeit sollen deshalb genau diese Aspekte aufgezeigt werden. Es werden Fragen zu Sicherheitskonzepten und deren Implementationen im Rahmen von GSM und GPRS gestellt und es werden Schwachstellen und mögliche Massnahmen zu deren Beseitigung aufgezeigt.*

## Inhaltsverzeichnis

## 9.1 Einleitung

Im Mobilfunkbereich bewegen wir uns in einer Welt, die mehr als eine Milliarde Menschen weltweit betrifft. [5] Die angebotenen Dienste sind aus unserem Leben kaum mehr wegzudenken und in unseren Breitengraden längst zu einer Alltäglichkeit geworden. Für viele kritische Abläufe ist das Mobilfunknetz von zentraler Bedeutung. Insbesondere im Geschäftsumfeld, wo man sich zunehmend mobilen Zugriff auf vertrauliche und sensitive Daten wünscht, ist deshalb die Frage nach der Sicherheit dieser Dienste zentral.

Die vorliegende Arbeit beschäftigt sich deshalb genau mit dieser Frage. Es geht um die Sicherheitseigenschaften vom „Global System for Mobile Communication“ (GSM) und dessen aufgesetztem Datenservice „General Packet Radio Service“ (GPRS). Die Arbeit ist im Rahmen eines Seminars zum Thema „Mobile Systems“ des Instituts für Informatik der Universität Zürich entstanden. Das Ziel der Arbeit ist es, einen Überblick über vorhandene Spezifikationen und Implementationen im Sicherheitsbereich zu geben, aber auch diese Mechanismen kritisch zu hinterfragen, um so ein Gesamtbild der Sicherheit im Mobiltelefonbereich zu erhalten. Ein weiteres Ziel ist es, den Leser auf sicherheitsspezifische Fragen zu sensibilisieren. Auf WAP-Services und Short Message Service (SMS) wird in dieser Arbeit aus Gründen des Umfangs nicht eingegangen.

Um die oben erwähnten Ziele zu erreichen, werden zuerst die Sicherheitsmechanismen anhand der einzelnen Architekturen aufgezeigt. In einem weiteren Schritt werden diese Mechanismen mittels den relevanten Sicherheitsaspekten analysiert und mögliche Lösungsansätze aufgezeigt.

## 9.2 GSM – Überblick

Die mobile Telekommunikation ist aus unserer heutigen Gesellschaft nicht mehr wegzudenken. Sie ist ein riesiges, weltweites Geschäftsfeld geworden und bedient momentan über eine Milliarde Kunden [5]. Der mit Abstand grösste Teil des Kuchens wird mit dem Standard „Global System for Mobile Communications“, kurz GSM, abgedeckt. Im Folgenden werden die Entstehung, der grundlegende technische Aufbau eines GSM- Systems, sowie die wichtigsten dabei verwendeten Technologien beschrieben. Da eine detaillierte Darstellung der oben erwähnten Punkte mit Leichtigkeit mehrere Bücher füllen könnte, beschränkt sich diese Seminarausarbeitung auf die Erläuterung der Hauptpunkte von GSM ohne Anspruch auf Vollständigkeit.

### 9.2.1 Entstehung von GSM

Die Entstehungsgeschichte von GSM geht auf das Jahr 1982 zurück. Das Ziel war es für Europa einen Standard für mobile Telekommunikation zu definieren um mehrere proprietäre Alleingänge einzelner Länder zu verhindern, die zu nicht kompatiblen Teilsystemen verteilt über ganz Europa geführt hätten, wie dies bei früheren analogen Systemen der Fall war. Zu diesem Zweck wurde Die „Group Spéciale Mobile“ am Europäischen Institut

für Telekommunikationsnormen (ETSI) eingerichtet, wo fortan die Spezifikationen für den GSM Standard erarbeitet wurden. Seither wurden mehrere GSM Standards spezifiziert. Um 1992 wurden in Europa dann die ersten kommerziellen Netze in Betrieb genommen und der weltweite Siegeszug von GSM ist bis heute ungebrochen.

### 9.2.2 Grundlagen von GSM

Wie bereits erwähnt, wurden verschiedene GSM Standards spezifiziert, welche jeweils einem anderen Frequenzband zugeordnet sind. So benutzt zum Beispiel GSM 1800 das Frequenzband von 1710 bis 1880 MHz. Der grundsätzliche Aufbau ist jedoch bei allen GSM Standards derselbe. GSM Systeme bieten die Möglichkeit eine zellenbasierte Antennenabdeckung aufzubauen, in welchem die Benutzer mit mobilen Geräten untereinander, aber auch mit netzfremden Teilnehmern und mit Festnetzanschlüssen, kommunizieren können. GSM bietet Techniken, die den Hauptschwierigkeiten der mobilen Telekommunikation gerecht werden. Diese sind die (genügend) gute Signalübertragung über die Luftschnittstelle, damit die Daten überhaupt verarbeitbar sind, sowie das Problem der Übergabe eines Mobilgerätes, sei dies von der einen zur anderen Zelle innerhalb desselben Netzbetreibers (Handover) oder die Nutzung eines fremden Mobilfunknetzes (Roaming).

### 9.2.3 Signalübertragung in GSM

Um digitale Signale, seien dies Daten oder Sprache, über die Luft zu übertragen, müssen die Zahlenströme bestehend aus Einsen und Nullen in Schwingungen umgesetzt werden. Dieser Vorgang wird als Modulation bezeichnet. GSM verwendet ein Variante der Frequenzmodulation, das „Gaussian Minimum Shift Keying“ [1]. Das nach der Modulation bestehende Signal ist aber für die direkte Übertragung ungeeignet, so dass es auf eine Trägerfrequenz aufgesetzt wird, welche besser geeignete Übertragungseigenschaften aufweist. Bei den meistgenutzten GSM Netzen sind es die Frequenzen im den Bereichen von 900, 1800 und 1900 MHz. In Europa sind hauptsächlich die beiden erstgenannten Frequenzbänder im Einsatz, während letzteres vor allem in Nordamerika benutzt wird. Da nun nicht alle Benutzer gleichzeitig dieselbe Frequenz belegen können, sind zur Kapazitätserhöhung drei verschiedene Multiplexverfahren in GSM umgesetzt.

Das erste und wohl simpelste Verfahren ist das Raummultiplexing. Es bedeutet, dass die gleiche Frequenz in einem Netzwerk von mehreren Zellen (d.h. die Reichweite einer einzelnen Antenne) genutzt werden kann, sofern sich die Zellen nicht überschneiden. Die zweite Multiplexvariante ist das Frequenzmultiplexing. In GSM wird das gesamte zur Verfügung stehende Frequenzspektrum in 200 KHz grosse Kanäle zerteilt, die nun alle zur selben Zeit für Übertragungen benutzt werden können. Eine einzelne Verbindung benützt jedoch zwei dieser Kanäle, einen für die Uplinkverbindung und einen für die Downlinkverbindung. Man spricht in einem solchen Fall von einem Frequenzduplex. Durch diese Einteilung stehen zum Beispiel in GSM 1800 374 Duplexkanäle zur Verfügung. Um die Kapazität noch einmal zu Erhöhen wird auch ein Zeitmultiplexverfahren eingesetzt. Jeder einzelne 200 KHz Kanal wird in Rahmen unterteilt, der jeweils acht Zeitschlitze von 577 Mikrosekunden. In diesem Zeitschlitz befinden sich nun die Nutzdaten, aber

auch Kontrolldaten zur Empfangsverbesserung, sowie ein Schutzabstand zu den folgenden Zeitschlitzen. Mit dieser Technik sind in unserm Beispiel GSM 1800 also 374 mal 8, gleich 2992 logische Kanäle gleichzeitig möglich.

### 9.2.4 GSM Architektur

#### Benutzer

Der Benutzer benötigt zwei Dinge, damit er die Möglichkeit hat ein Mobilnetz eines bestimmten Anbieters zu Nutzen. Erstens muss er im Besitz eines GSM-fähigen Mobiltelefons sein. Dabei spielt der Hersteller keine Rolle, sofern die Spezifikationen eingehalten sind. Jedes Mobiltelefon ist mit einer eindeutigen Nummer der „International Mobile Equipment Identity“ (IMEI). Die zweite benötigte Komponente ist das „Subscriber Identity Module“, hierzulande meist SIM-Karte genannt. Diese SIM-Karte identifiziert den Benutzer eindeutig gegenüber dem Betreiber mittels der „International Mobile Subscriber Identity“ (IMSI). Des Weiteren enthält sie Daten wie PIN-Code, Authentifizierungsschlüssel, Seriennummer und andere benutzerspezifische Daten [4].

#### Netzbetreiber

Aus Sicht eines Betreibers des GSM Netzes sieht die ganze Sache nicht mehr so einfach aus. Grundsätzlich kann man aber drei Subsysteme unterscheiden, das Funk-Festnetzsystem (Radio Subsystem), das Mobilvermittlungssystem (Network and Switching Subsystem) und das Betriebs- und Wartungssystem (Operation Subsystem) [4]. An dieser Stelle wird nur auf die für die Sicherheit relevanten Teilkomponenten des Systems eingegangen (Siehe Abbildung 9.1). Wer eine komplette Beschreibung der Architektur wünscht, wendet sich bitte an die entsprechende Fachliteratur.

Die wohl wichtigste Komponente für Sicherheitsaspekte ist das „Home Location Register“ (HLR). Dieser Speicher enthält alle benutzerspezifischen Daten wie Telefonnummern, „International Mobile Subscriber Identity“ (IMSI) oder den momentan gültigen Aufenthaltsort des Benutzers. Dem HLR angeschlossen ist das „Authentication Center“ (AuC). Hier werden die besonders schützenswerten persönlichen Schlüssel der Kunden gespeichert und alle Berechnungen, die mit der Authentifizierung oder Verschlüsselung in Zusammenhang stehen, durchgeführt. Eine ähnliche Funktion wie der HLR übernehmen die „Visitor Location Register“ (VLR). Auch hier werden benutzerspezifische Daten gespeichert, jedoch nur solche die sich auf Benutzer beziehen, die sich in dem Bereich befinden für den der VLR zuständig ist. Die VLRs sind sozusagen Entlastungsstellen für den HLR, damit dieser nicht alle Anfragen bearbeiten muss [4].

Die letzte Komponente, die für sicherheitsspezifische Operationen genutzt werden kann, ist der „Equipment Identity Register“ (EIR). In dieser Datenbank werden alle IMEIs, also alle Identifikationsnummern der Telefone (also nicht die SIM Identifikation) gespeichert. Diese Liste verhindert, dass gestohlene mobile Geräte mit regulären SIM-Karten genutzt werden können, da der Netzbetreiber in der Lage ist die betreffenden IMEIs zu sperren

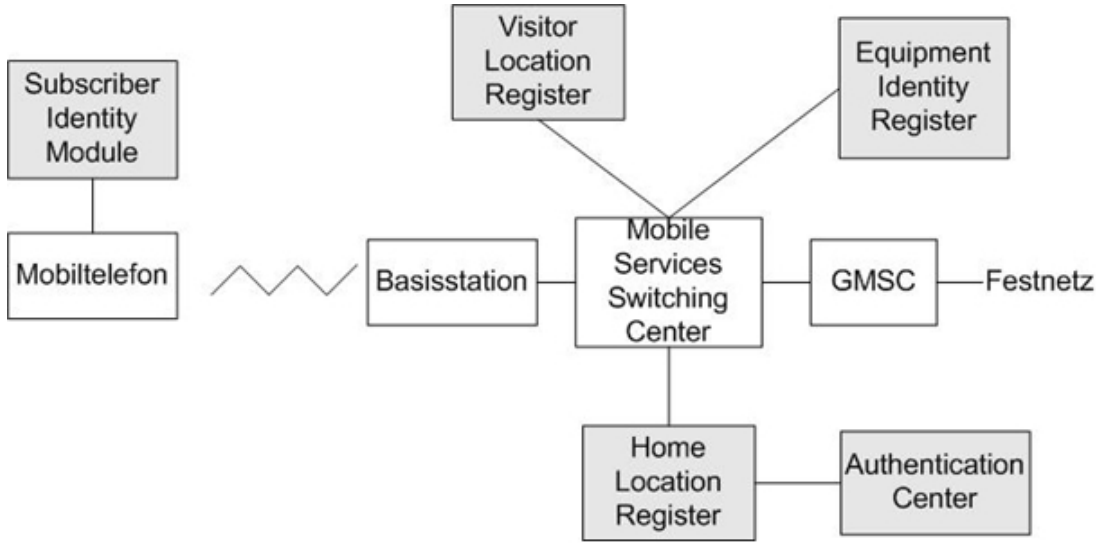


Abbildung 9.1: Vereinfachte Darstellung der GSM Architektur, mit Blickpunkt auf die sicherheitsrelevanten Komponenten [Eigene Darstellung]

und sogar den jetzigen Benutzer zu eruieren. Der EIR enthält eine weisse, eine graue und eine schwarze Liste. Auf der weissen sind alle zugelassenen IMEIs, auf der schwarzen alle gesperrten und auf der grauen alle IMEIs deren Aufenthalt und Benutzung aufgezeichnet wird [1]. Eine Sperrung und somit eine Zuweisung auf die schwarze Liste erfolgt einerseits für als gestohlen gemeldete mobile Geräte, aber auch für Telefone, welche aufgrund von Fehlfunktionen oder Fehlmanipulationen die Sicherheit des GSM Netzes gefährden und deshalb von diesem Ausgeschlossen werden.

## 9.3 Sicherheit in GSM

Nach der Darlegung der grundlegenden Eigenschaften und den benutzten Technologien eines GSM Netzes kommt nun der Punkt auf den das Hauptaugenmerk dieses Papers gerichtet ist, auf die Sicherheit in GSM. Es werden vier sicherheitsrelevante Aspekte unterschieden, welche in einem GSM Netz abgedeckt sein müssen. Es sind dies Anonymität, Authentifizierung, Schutz der Signalisierungs- und der Benutzerdaten [2]. Im Folgenden werden diese Eigenschaften kurz beschrieben, gefolgt von deren Umsetzung in GSM und der den Vorgängen zu Grunde liegenden Algorithmen.

### 9.3.1 Sicherheitsaspekte in GSM

Die Spezifikationen von GSM sehen drei Teilaspekte vor, welche im Bereich Sicherheit abgedeckt werden müssen. Zwei der Bereiche zielen auf die Sicherheit des Benutzers des GSM Netzes, während Aspekt Nummer drei den Netzbetreiber vor unlauteren Machenschaften schützen soll [2]:

- Anonymität

  Ein GSM Netz soll dem einzelnen Mobiltelefonbenutzer eine gewisse Anonymität und Schutz der Privatsphäre zusichern. Das heisst seine Identität soll, ausser dem Gesprächsteilnehmer und der Telefongesellschaft für Rechnungszwecke, niemandem zugänglich sein.

- Authentifizierung

  Bei dieser Sicherheitsüberlegung handelt es sich um den Schutz des GSM Netzanbieters vor betrügerischen Manipulationen auf seine Kosten. Eine klare und eindeutige Benutzerauthentifizierung verhindert solches und garantiert auch dem Benutzer eine korrekte Abrechnung seiner beanspruchten Dienste.

- Schutz der Signale und der Benutzerdaten (Verschlüsselung)

  Der dritte Aspekt betrifft die Übermittlung der Signalisierungs- und Benutzerdaten über die Luftschnittstelle. Da das Medium Luft nicht wie eine Kabelverbindung gegen fremden (physischen) Zugriff gesichert werden kann, müssen die Daten verschlüsselt gesendet werden. Die gilt für die Gesprächsdaten wie auch für sensitive Signalisierungsdaten (z.B. Telefonnummer).

### 9.3.2 Umsetzungen der Sicherheitsaspekte in GSM

Nachdem die für GSM vorgesehenen Sicherheitsaspekte kurz vorgestellt worden sind, richten wir unseren Blick auf das Wesentliche, nämlich wie diese Bereiche für GSM spezifiziert worden sind und wie die konkrete Umsetzung dieser Spezifikationen der drei Teilaspekte der Sicherheit aussieht.

#### Anonymitätsmechanismen

Wie weiter oben erwähnt, besitzt jedes „Subscriber Identity Module“ eine eindeutige Identifizierung; die IMSI (International Mobile Subscriber Identity). Durch diese unverschlüsselte Identifizierung kann bei einem Lauschangriff darauf geschlossen werden, welche Person sich momentan im Sendebereich der abgehörten Basisstation befindet. Um die Anonymität der Kunden vor solch simplen Angriffen zu schützen, setzt man in GSM auf temporäre Identifizierung mittels TMSI (Temporary Mobile Subscriber Identity).

Während die IMSI nur bei der ersten Anmeldung eines mobilen Gerätes ans GSM Netz benutzt wird, erhält der Benutzer danach eine TMSI, welche nur noch das GSM System die wahre Identität hinter dieser Identifikation ermitteln lässt. Die aktuelle TMSI wird innerhalb des Netzwerkes zur IMSI im aktuellen VLR (Visitor Location Register) abgespeichert und somit bleiben die benutzten Dienste jederzeit exakt zurechenbar. Gleichzeitig wird die TMSI auch auf dem SIM des Benutzers abgespeichert und bei der nächsten Anmeldung ans GSM Netz wieder verwendet. Die TMSI wird mehrmals geändert, zum Beispiel, wenn ein Handover durchgeführt werden soll, oder aber auch nach einer festgelegten Zeitspanne. Der Änderungsvorgang der TMSI wird nun aber verschlüsselt durchgeführt. Somit kann eine Verfolgung der TMSI und damit das erkennen der Benutzeridentität ausgeschlossen werden [3].

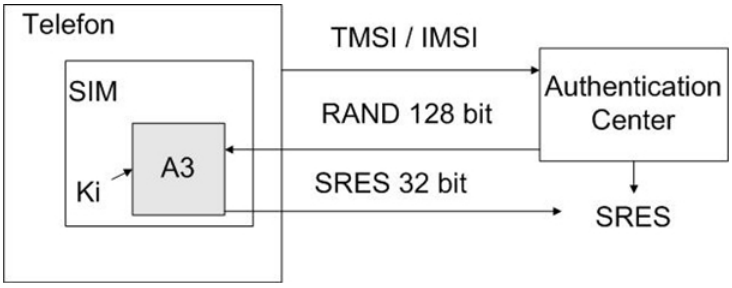


Abbildung 9.2: Authentifizierung mit GSM [Eigene Darstellung]

### Der Authentifizierungsvorgang in GSM

Die zweifelsfreie und eindeutige Authentifizierung ist ein Vorgang der nun nicht mehr ganz so einfach umzusetzen ist wie noch die Anonymität. Der in GSM benutzte Ansatz ist das „Challenge and Response"-Verfahren (Abbildung 9.2). Die erste Voraussetzung für eine erfolgreiche Anmeldung am System ist das Bekanntsein eines für jeden Benutzer eindeutigen Schlüssels Ki. Dieser 128 bit lange Schlüssel ist einerseits auf dem SIM und anderseits im AuC (Authentification Center) des Netzwerkes gespeichert [4]. Um diesen Schlüssel vor Missbrauch zu schützen, wird er niemals über die Luftschnittstelle übertragen.

Die zweite Komponente, die für das „Challenge and Response"-Verfahren benötigt wird, ist ein Algorithmus, der die entsprechenden Verschlüsselungen durchführt. In der GSM Spezifikation ist dies der Algorithmus A3. Dies ist jedoch keine Bezeichnung für einen einzelnen Algorithmus, sondern eine Bezeichnung für alle eingesetzten Authentifizierungsalgorithmen in GSM [3]. In den meisten GSM Netzen ist A3 als eine von zwei COMP128 Algorithmus Versionen implementiert [6], theoretisch ist auch eine andere Algorithmuswahl für die Umsetzung von A3 möglich. Wieso gerade dieser COMP128 Algorithmus in Einsatz ist, wird im Abschnitt über die Probleme mit den Algorithmen eingegangen.

Nun sind die Voraussetzungen für den Authentifizierungsvorgang bekannt, jetzt kommt also die konkreten Durchführung. Das mobile Gerät nimmt Verbindung mit dem Netzwerk auf und sendet entweder seine TMSI oder die IMSI zur Identifizierung. Das Netzwerk sendet dem Anfrager eine 128 bit lange Zufallszahl RAND zu. Im SIM des mobilen Gerätes wird nun diese Zufallszahl RAND zusammen mit dem Schlüssel Ki durch den Algorithmus A3 verschlüsselt und eine 32 bit lange gekennzeichnete Antwort (SRES, Signed Response) wird erzeugt. Diese SRES sendet das mobile Gerät ans Netzwerk zurück. Diese SRES des Telefons wird nun mit der eigenen SRES des Systems verglichen. Diese wurde im AuC mittels des dort für die betreffende IMSI oder TMSI hinterlegten Schlüssels Ki erstellt. Falls die beiden SRES übereinstimmen darf das Telefon das Netz benutzen. Sollte der Vorgang fehlschlagen, wird, falls eine TMSI verwendet wurde, die Authentifizierung mit der IMSI wiederholt oder, bei IMSI Verwendung, die Verbindung abgebrochen.

### Verschlüsselung in GSM

Der dritte Aspekt der Sicherheit, die geschützte Übertragung von Daten über die Luftschnittstelle, erreicht man durch deren Verschlüsselung. Zwar sind die gesendeten Infor-

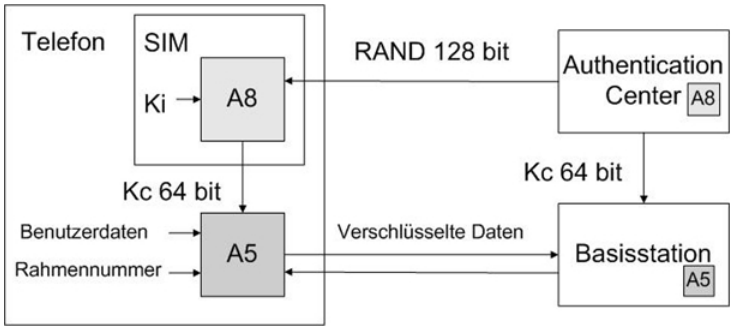


Abbildung 9.3: Verschlüsselung mit GSM [Eigene Dartstellung]

mationen immer noch abhörbar, jedoch ohne den entsprechenden Dechiffrierungsschlüssel völlig nutzlos für den Angreifer. Um eine gute Verschlüsselung zu erreichen verwendet GSM zwei Algorithmen, einen zum erzeugen des Chiffrierungsschlüssels Kc, den zweiten zum eigentlichen Verschlüsseln der Nachricht (Siehe Abbildung 9.3).

Der Verschlüsselungsvorgang beginnt eigentlich schon mit der Authentifizierung. Sobald das mobile Gerät die Zufallszahl RAND erhält, wird auf dem SIM der Algorithmus A8 mit den Inputs RAND und den persönlichen Schlüssel Ki durchgeführt. Das Ergebnis dieses Vorgangs ist der 64 bit lange Chiffrierungsschlüssel Kc. Derselbe Ablauf geschieht im AuC des GSM Netzes, worauf der Schlüssel Kc an die für die Verbindung zuständige Basisstation weitergeleitet wird (respektive an die entsprechende BTS (Base Transceiver Station)). Der Algorithmus A8 ist ebenfalls wie A3 nicht festgelegt, das heisst, er kann vom Provider selbst implementiert werden. Meistens ist aber auch dieser Algorithmus eine Version von COMP128 und wird gemeinsam mit A3 auf dem SIM implementiert (benannt als A3/A8) [5].

Nun besitzen der Benutzer und die zuständige Basisstation einen eindeutigen Schlüssel, den sie zur Chiffrierung dieser einzelnen Übertragung benutzen können. Zusätzlich zum Schlüssel wird die 22 bit lange TDMA-Rahmennummer (Länge jedes TDMA Rahmen: 4.615ms) als Input in den Algorithmus A5 verwendet. Der Output dieses Vorganges ist ein 114 bit langer binärer Datenstrom. Dieser wird mit den Nutzdaten geXORed. Der somit verschlüsselte Text wird nun an die Gegenstation übertragen. Dort sind ja Schlüssel und Rahmennummer ebenfalls bekannt, so dass der Datenstrom dechiffriert werden kann und die Nachricht verstanden wird.

Der Algorithmus A5 unterscheidet sich in wichtigen Eigenschaften von A3 und A8. Einerseits kann er nicht beliebig implementiert werden, sondern ist vorgegeben, damit die Interoperabilität der mobilen Geräte in GSM Netzen verschiedenster Anbieter gewährleistet wird. Andererseits wird der Algorithmus um eine hohe Rechengeschwindigkeit zu gewährleisten nicht im SIM, sondern im mobilen Gerät selbst, also als Hardwarebestandteil, implementiert. Im Moment existieren drei Versionen von A5.

### Die Algorithmen A3, A5 und A8

Dieser Abschnitt bietet einen etwas tieferen Einblick in die in einem GSM Netz zur Authentifizierung und Verschlüsselung verwendeten Algorithmen A3, A5 und A8. Wie bereits

erwähnt werden die Algorithmen A3 und A8 meist in ähnlicher Form umgesetzt und es gelten für beide die gleichen Voraussetzungen. Beide werden auf dem SIM implementiert und es gibt für den Betreiber keine Vorschriften wie die Algorithmen umzusetzen sind, da keine Interoperabilitätsprobleme auftreten können, weil im Falle von Roaming auf den HLR des Kartenausgebers zurückgegriffen wird [8]. Der einzige grosse Unterschied ist der Zweck der beiden Algorithmen. A3 hat Authentifizierungsaufgaben, indem er eine 32 bit lange SRES erzeugt, während A8 einen 64 bit Schlüssel für die Datenchiffrierung generiert.

Die wirkliche Realisierung der beiden Algorithmen war lange Zeit unbekannt, da nach der Philosophie SSecurity by Obscurity"gehandelt worden ist. Jedoch hat das GSM Entwicklungskonsortium seinen Mitgliedern (also den Mobilfunkbetreibern) eine Referenzimplementierung, oder besser gesagt eine Beispielimplementierung von A3 und A8, den Algorithmus COMP128 zur Verfügung gestellt, der in den meisten GSM Netzen umgesetzt wurde. Diese wurde aber 1998 durch Reverse Engineering öffentlich gemacht [9]. Es existieren drei Versionen von COMP128, die sich bezüglich ihrer Sicherheit unterscheiden. Auf diese Mängel wird im Kapitel Algorithmenprobleme eingegangen.

Beim Verschlüsselungsalgorithmus A5 präsentiert sich die ganze Situation ein bisschen anders. Da die Verschlüsselung mit allen Mobiltelefonen auch in Netzten von fremden Dienstanbietern funktionieren muss, ist in allen Telefonen derselbe Algorithmus einzusetzen. Auch hier gibt es verschiedene Versionen, das Protokoll lässt bis zu sieben Versionen zu. Am meisten verbreitet sind die Versionen A5/1, welcher nur von Mitglieder von CEPT (The European Conference of Postal and Telecommunications Administrations) benutzt werden darf und die Version A5/2, eine schwächere Umsetzung, welche in allen anderen Ländern zum Einsatz kommt [2]. Beide Versionen basieren auf der Verschaltung von linear rückgekoppelten Schieberegistern (LFSR) [10]. Eine dritte Bezeichnung A5/0 wird für unverschlüsselte Datenübertragung verwendet. Ebenfalls wie die beiden andern Algorithmen wurde A5 nicht offen gelegt. Jedoch sind die verschiedenen Versionen von A5 durch mehrere Veröffentlichungen im Internet mittlerweile bekannt.

### 9.3.3 Sicherheitsrelevante Probleme in der GSM Architektur

Obwohl bereits in der Spezifizierungsphase von GSM klare Sicherheitsvorgaben erlassen wurden und diese Sicherheitsdienste in der Architektur und von den Netzbetreibern umgesetzt worden sind, kann GSM nicht als hundertprozentig sicher betrachtet werden. Die folgenden Abschnitte beleuchten fünf sicherheitsrelevante Fragestellungen, Probleme und gar Mängel des GSM Systems. Diese sind die Einweg- Authentifizierung, das Anonymitätsproblem, die Schwächen der verwendeten Algorithmen, die Probleme mit der Umsetzung der EIR und zu guter letzt die Ende-zu-Ende Kommunikation. Einige dieser Probleme sind eng miteinander verknüpft oder entstehen erst auseinander (wie das Anonymitätsproblem aus der Einweg-Authentifizierung), andere sind eher als Hinweise zu sehen auf Bereiche, wo bessere Sicherheit ohne riesigen Mehraufwand relativ leicht zu gewinnen wäre (EIR-Umsetzung, Ende-zu-Ende Kommunikation). Falls für die jeweiligen Probleme bereits Lösungen vorliegen oder Lösungsansätze diskutiert werden, so wird deren Betrachtung im selben Kapitel abgehandelt.

**Problem der Einweg-Authentifizierung**

Das wohl grösste Problem in der GSM Sicherheitsarchitektur, welches viele der nachfolgenden Probleme erst entstehend lässt, ist die Einweg-Authentifizierung. Dies bedeutet, dass sich das Mobiltelefon zwar eindeutig beim Netzwerk anmelden und identifizieren muss, in umgedrehter Richtung jedoch nichts geschieht. Das heisst die Basisstation muss dem mobilen Gerät nicht beweisen, dass sie den richtigen Schlüssel Ki kennt [3]. Das Fehlen eines beidseitigen Authentifizierungsvorganges in GSM Netzen führt nun zu einer Möglichkeit für Angreifer die gesamten Sicherheitsvorkehrungen zu brechen. Der Angreifer kann eine eigene Basisstation aufstellen und mit dieser mit Mobiltelefonen kommunizieren, ohne dass diese einen Unterschied zum richtigen GSM Netz bemerken. Das Aufsetzen einer falschen Basisstation kann dazu benutzt werden um den persönlichen Schlüssel Ki eines Benutzers zu eruieren, ohne physischen Kontakt zu dessen Telefon oder, genauer gesagt, zu dessen SIM zu haben.

Eine andere Möglichkeit wäre das Benutzen einer falschen Basisstation um Lauschangriffe zu starten: Alle über diese Basisstation eingehenden Anrufe oder Datenströme werden abgehört, dann aber wie bei einer guten Basisstation weitergeleitet („Man-in-the-middle"-Angriff) [3]. Dazu wird eine vollständig funktionierende Basisstation mit voller Funktionalität benötigt. Diese agiert wie eine normale Basisstation des Providers, dass heisst die eingehenden Anrufe werden korrekt ins GSM Netz weitergeleitet. Der Unterschied besteht darin, dass alle Verschlüsselungsmechanismen deaktiviert sind, so dass alle Gespräche und Daten die über seine Basisstation laufen für den Angreifer völlig transparent sind.

Auf den ersten Blick scheinen diese Angriffsmöglichkeiten wegen den technischen und finanziellen Aufwendungen eher theoretischer Natur zu sein. Vor allem eine vollständig integrierte Basisstation, die die korrekte Weiterleitung verwaltet, welche für den Middle-Angriff nötig ist, scheint unrealistisch. Jedoch eine einfache, abgespeckte Version einer Basisstation, welche nur auf das Brechen des Schlüssels Ki ausgelegt ist, lässt sich schon für etwa 10'000 Dollar verwirklichen [14]. Dies ist zwar immer noch nicht wenig, doch scheint eine Verwirklichung einer falschen Basisstation in diesem finanziellen Rahmen viel weniger unrealistisch. Die vereinfachte Version kann nun all Telefone, die versuchen sich über diese Station ins GSM Netz anzumelden gefangen halten, indem unzählige Authentifizierungsversuche durchgeführt werden. Im Laufe dieser Versuche ist es möglich den Authentifizierungsschlüssel Ki ohne physischen Zugriff auf das SIJM zu ermitteln.

**Anonymitätsproblem**

Die Anonymität der Benutzer eines GSM Netzes wird, wie bereits beschrieben, durch den Einsatz von temporären Identifizierungsangaben (TMSI) gewährleistet. Doch der vorgestellte Einsatz einer falschen Basisstation bringt das gesamte Konstrukt zum Einsturz. Das GSM Netz verwaltet alle momentanen TMSIs in den VLRs. Sollte nun einmal, aus welchen Gründen auch immer, eine TMSI verloren gehen oder nicht mehr bekannt sein, so hat das Netzwerk, respektive die Basisstation, in dessen Zelle sich das mobile Gerät mit der ungültigen TMSI befindet, die Möglichkeit die IMSI des SIM anzufordern. Dieser Identifizierungsaufforderung hat nun das Mobiltelefon auf jeden Fall nachzukommen und

sendet die IMSI unverschlüsselt zum Anfrager [3]. Durch diesen Vorgang ist es nun für falsche Basisstationen ein Leichtes die wahren Identitäten der Benutzer im Einzugsgebiet zu ermitteln.

### Algorithmenproblem

Ein Problem, der mehr als Fehler der Spezifizierungsgruppe betrachtet werden muss, ist das Handeln nach „Security by Obscurity“, dass heisst die für Authentifizierung und Verschlüsselung verwendeten Algorithmen A3, A5 und A8 wurden nicht offen gelegt. Folglich konnten somit nicht von der breiten (Informatik-)Öffentlichkeit getestet und geprüft werden. Und so kam es wie es kommen musste. Um 1998 tauchten die verwendeten Implementierungen im Internet auf und schon bald stiess man auf Sicherheitsmängel und Möglichkeiten die der Verschlüsselung zu Grunde liegenden Schlüssel zu knacken.

Die Algorithmen A3 und A8 sind, wie bereits erwähnt, gemeinsam als COMP128 Algorithmus auf dem SIM implementiert. COMP128 war jedoch eigentlich nur als eine Beispielimplementierung herausgegeben worden und trotzdem wurden sie von den meisten Anbietern ohne grosse Änderungen oder Prüfungen übernommen. Wegen der schlechten Rechenleistung der SIM werden für den Algorithmus vorausberechnete Tabellen verwendet. Dies führt zu physikalisch messbaren Mustern. Somit stehen Möglichkeiten für einen Side Channel- Angriff zur Verfügung um an den geheimen Schlüssel zu gelangen [13]. Eine weiter Möglichkeit, um an den persönlichen Schlüssel Ki des Benutzers zu gelangen, ist die gezielte Wahl des Wertes RAND. Eine Schwäche im COMP128 Algorithmus führt dazu, dass die Anzahl der Versuche um an den Schlüssel zu gelangen von bestenfalls 2128 auf 213 bis 215 (= 7000 bis 35000) gesenkt werden kann. Diese Art von Angriff genötigt zwar physischen Zugang zum SIM, doch ist dieser gegeben, lässt sich der Schlüssel Ki innert einer Stunde extrahieren [3].

Eine weitere, sehr gewichtige Schwäche des COMP128 Algorithmus ist die Umsetzung des Sitzungsschlüsselgenerators A8. Zwar wird jeweils ein 64 bit langer Schlüssel Kc generiert, doch werden 10 bit davon auf Null gesetzt. Dies reduziert demzufolge die Sicherheit des Chiffrierungsschlüssels von 64 auf 54 bit [4]. Dieses Problem, welches in den Versionen eins und zwei des COMP128 Algorithmus auftritt, verringert die Sicherheit des Schlüssels Kc um den Faktor 1024 [3]. Glücklicherweise wurde dieses Problem ernst genommen und in der dritten Ausführung des Algorithmus, COMP128-3 soll diese Lücke geschlossen worden sein [15]. Jedoch ist anzunehmen, dass immer noch eine Mehrzahl der benutzten SIM mit den kompromittierten Algorithmen arbeitet.

Nun zum Algorithmus A5, der für die eigentliche Datenverschlüsselung zuständig ist. Da dieser Vorgang netzübergreifend verfügbar sein muss, ist die Implementierung vorgegeben. Von Anfang an wurden zwei Versionen herausgegeben: Version A5/1 für Mitglieder von CEPT und eine bereits geschwächte Version A5/2 für alle anderen Länder. Auch diese beiden Implementierungen wurden nach der Vorgehensweise „Security by Obscurity“ geheim gehalten, jedoch wurden sie ebenso wie COMP128 durch teilweise Veröffentlichungen im Internet und durch Reverse Engineering offen gelegt [6]. So wurden auch bald darauf gravierende Mängel entdeckt, welche die Verschlüsselung relativ schnell brechen lassen.

So wurden zwei verschiedene Angriffsmuster entdeckt, mit deren Hilfe der Schlüssel mittels eines einzigen PCs ermittelt werden kann. Beide Varianten benötigen eine grosse Vorberechnungszeit von 248 Schritten, welche allerdings nur einmal und nicht für jeden einzelnen Angriff, durchgeführt werden muss. Danach benötigt die erste Variante zwei Minuten der mit A5/1 verschlüsselten Daten während eines Gesprächs, um dann innert einer Sekunde den Schlüssel Kc zu ermitteln. Für die Variante Zwei genügen zwei Sekunden des Gespräches um den Schlüssel in wenigen Minuten zu berechnen [11].

Somit ist die Wirksamkeit des Verschlüsselungsalgorithmus A5/1 mehr als in Frage gestellt, ganz zu schweigen von der schwächeren Version A5/2. Mit den entsprechenden Kenntnissen ist jedes verschlüsselte Gespräch, das belauscht wird, in wenigen Minuten abhörbar. Doch auch bei diesem Teil des GSM Sicherheitssystems ist eine mögliche Lösung in Sicht.

2002 wurde neue Version des A5 Algorithmus herausgegeben, der entgegen bisherigen Gepflogenheiten öffentlich zugänglich ist [12]. Diese Version A5/3 basiert auf dem KASUMI Algorithmus, der für die Verschlüsselung bei UMTS eingesetzt wird und bietet eine markant bessere Sicherheit als seine Vorgänger [10]. Nun liegt die verbesserte Sicherheit in den Händen der Gerätehersteller und bei den Netzbetreibern, welche angehalten sind diese neue Version in ihren Telefonen, respektive in ihren Basisstationen, umzusetzen und somit diese bestehende Sicherheitslücke endgültig zu schliessen.

## EIR Anwendung bei Providern

Ein etwas andere Aspekt der Sicherheitsfrage in GSM Netzen ergibt sich bei der Betrachtung einer speziellen Komponente der Gesamtarchitektur, des Equipment Identity Register (EIR). Der EIR enthält alle Gerätekennungen (IMEI) der Telefone, welche im Netz des Providers benutzt werden. Zu jeder IMEI wird die entsprechende Benutzeridentifikation (IMSI) festgehalten, so dass die Geräte eindeutig dem Benutzer momentanen Benutzer zugerechnet werden können. Diese Komponente ermöglicht nun auf einfache Art und Weise das Aufspüren von gestohlenen Mobiltelefonen. Das heisst, sowohl der ungefähre Aufenthaltsort des Benutzer, also in der Umgebung welcher Basisstation das Telefon benutzt wird, sowie auch die Identität des jetzigen Benutzers, sofern es sich nicht um einen anonymen Prepaid-Abonnementen handelt, können ermittelt werden.

Die eben erwähnten Eigenschaften sollten dem EIR einen wichtigen Platz im GSM Sicherheitskonzept einräumen. Doch dem ist nicht so. Das erste Problem sind die tiefen und immer noch sinkenden Kosten für mobile Geräte. Dies führt dazu, dass viele Provider eine Installation des EIR verschoben haben oder gar ganz davon absehen [1]. Des Weiteren ist keine einheitliche Spezifikation vorhanden, welche die Abfragen zum EIR regelt. Das aber weitaus gewichtigere Problem stellen die verschiedenen Provider, oder vielmehr deren jeweiligen EIR dar: Jeder Netzbetreiber verwaltet eine Datenbank mit den als gestohlen gemeldeten Mobiltelefonen. Somit werden die gestohlenen Geräte erkannt, wenn sie nochmals im selben GSM Netz verwendet werden. Da aber die Daten des EIR unter den verschiedenen Providern nur mangelhaft abgeglichen werden, ist die Benutzung eines gestohlenen Telefons im Netz eines fremden GSM Anbieters oft immer noch möglich [4].

So unschön diese Problematik auf den ersten Blick erscheint, zeichnet sich doch eine mögliche Verbesserung, ja vielleicht sogar Lösung des Problems ab: Im Jahre 2002 wurde der Central Equipment Identify Register (CEIR) in Betrieb genommen. Dieses zentrale Register soll die Daten der EIR aller Netzbetreiber vereinen, indem die jeweiligen EIR direkt auf diese totale Sammlung zugreifen können. Der CEIR könnte alle Datenabgleichprobleme auf einen Schlag lösen und somit den Einsatz von gestohlenen oder sonst nicht mehr zugelassenen Telefonen auf einen Schlag unterbinden. Leider jedoch sind momentan nur wenige Provider beim CEIR angeschlossen, so dass dessen Wirkungsgrad sehr beschränkt bleibt [5].

### Problem Ende-zu-Ende Kommunikation

Die in diesem Seminarbericht vorgestellten Sicherheitsmechanismen von GSM zielen ausnahmslos auf die Sicherheit der Daten während des Übermittelns über die Luftschnittstelle, also vom mobilen Gerät bis zur Basisstation. Wie in anderen Teilen aufgezeigt wurde, ist diese Teilstrecke aber alles andere als vollständig sicher; die Schwäche der Algorithmen oder auch grundlegende Fehler in der Sicherheitsarchitektur (Einweg-Authentifizierung) bieten potentiellen Angreifern eine grosse Auswahl von Angriffsmöglichkeiten.

Bis anhin wurde jedoch eine andere grosse Sicherheitslücke komplett ausser Acht gelassen. Wie beschrieben werden die Sprach- und Nutzdaten zwar im Mobiltelefon verschlüsselt und so zur Basisstation gesandt, doch an dieser Stelle erfolgt bereits die vollständige Entschlüsselung der Daten. Somit werden diese von der Basisstation des GSM Netzes bis zum Bestimmungspunkt, sei dies ein Festnetzanschluss oder eine andere Basisstation eines Mobilfunknetzes, ein leicht angreifbares Ziel. Natürlich gilt dies nur unter dem Gesichtspunkt, dass der Angreifer physischen Zugriff auf das unverschlüsselte Netzwerk erlangt. Sobald dieser Zugang aber besteht, sind die Daten einfach abzufangen und auszuwerten, da keinerlei Chiffrierung geknackt werden muss. Wenn der Benutzer eines GSM Netzes vollständige Sicherheit wünscht, dann muss er seinerseits etwas dafür unternehmen und auf Zusatzdienste, welche auf die bestehende GSM Architektur aufgesetzt werden, zurückgreifen.

Eine Möglichkeit um Datendienste, welche auf GSM basieren, zu schützen, ist der Einsatz eines Virtual Private Networks (VPN). Eine genauere Beschreibung der VPN befindet sich im Abschnitt über GPRS dieser Arbeit.

Es gibt aber auch eine Möglichkeit, um die sichere Ende-zu-Ende Übertragung von Sprachdaten zu gewährleisten, nämlich die „End-to-End Encryption“ (EEE, Ende-zu-Ende Verschlüsselung) [16]. Für diese Vorverschlüsselung der Sprachdaten, bevor sie den normalen GSM Chiffrierungsvorgang durch den Algorithmus A5 durchlaufen, wird eine spezielle Hardware benötigt. Ein Beispiel für ein Gerät, dass eine solche EEE besitzt ist das von der schwedischen Firma Sectra in Zusammenarbeit mit dem norwegischen Verteidigungsministerium entwickelte NSK 200. Dieses Telefon verfügt über eine solch gute EEE, dass es von der NATO sogar für die Sicherheitsstufe „NATO geheim“ zertifiziert wurde cite/sectra. Dieses Beispiel soll zeigen, dass durch die Benutzung eines Zusatzdienstes, beziehungsweise von Zusatzsoftware die Strecke vom Mobiltelefon bis zum Zielort des Anrufes vollständig

sicher gestaltet werden kann. Jedoch ist diese Sicherheit mit einem beträchtlichen Mehraufwand von Arbeit und/oder Kosten verbunden, so dass ein solcher Schutz der Leitung für den normalen GSM Benutzer unrealistisch ist.

**Beurteilung der Sicherheitsprobleme in GSM**

Die Spezifikation des "Global System for Mobile Communications"wurde vor über 20 Jahren begonnen und der weltumfassende Erfolg des Systems ist zweifellos. Doch die in diesem Paper erwähnten Sicherheitsprobleme zeigen klar und deutlich, dass das GSM System nicht als sichere Kommunikationsplattform betrachtet werden kann. Es wurden klare Spezifikationsfehler begangen, wie etwa bei der Unterlassung der beidseitigen Authentifizierung zwischen mobilem Gerät und der Basisstation. Dies führt nahtlos zu der darauf hinfälligen Netzanonymität des einzelnen Benutzers. Des Weiteren sind die eingesetzten Algorithmen so schwach, dass die gesamte Authentifizierungs- und Verschlüsselungsprozedur arg komprimittiert ist. Hinzu kommen der ungenügende Einsatz vorhandener Mittel (Equipment Identity Register Einsatz), sowie die mangelnde Sensibilität gegenüber der Nutzungssicherheit eines Mobilfunknetzes in der Gesamtsicht (Ende-zu-Ende Kommunikation). Zwar sind bei den meisten Teilproblemen mögliche Verbesserungen in Sicht, doch kommt deren Einsatz eher schleppend in Gang.

## 9.4 GPRS – Überblick

Neben der mobilen Sprachkommunikation mit GSM wurden auch Datendienste wie das Internet immer populärer. Bis zum Jahr 2000 blieben diese Dienste von der mobilen Welt praktisch ausgeschlossen, da GSM mit WAP bis dahin nur 9.6 kbit/s anbieten konnte (fünfmal langsamer als ein Standardanalogmodem). Aus dem Bedürfnis nach mobiler Datenkommunikation heraus, das vor allem aus der Businesswelt stammte, wurden neue Standards definiert. Neben der völlig neuen Technologie UMTS, das eine eigene Infrastruktur benötigt, wurde im Jahr 2000 der General Paket Radio Service, kurz GPRS, in Betrieb genommen.

Dieser Standard ermöglicht auf dem GSM Netz eine paketorientierte Datenübertragung mit bis zu 50 kbit/s. Mit dem paketorientierten GPRS ergaben sich für die Benutzer, wie auch für die Betreiber, einige wichtige Vorteile. Da dem Nutzer nun nicht mehr – unabhängig von der Datenmenge – eine ganze Leitung zugesichert werden muss, kann der Betreiber seine Leitungen viel effizienter nutzen und der Benutzer kann dauernd online sein und bezahlt nur die effektiv beanspruchte Datenmenge. [18] Wie wichtig die Datenservices sind, zeigen Zahlen von Swisscom Mobile für die Periode von 2000 bis 2003. In diesem Zeitraum, direkt nach Einführung von GPRS, hat sich die Datenmenge auf dem Netz des Betreibers von 5 auf über 12 Prozent mehr als verdoppelt. [19]

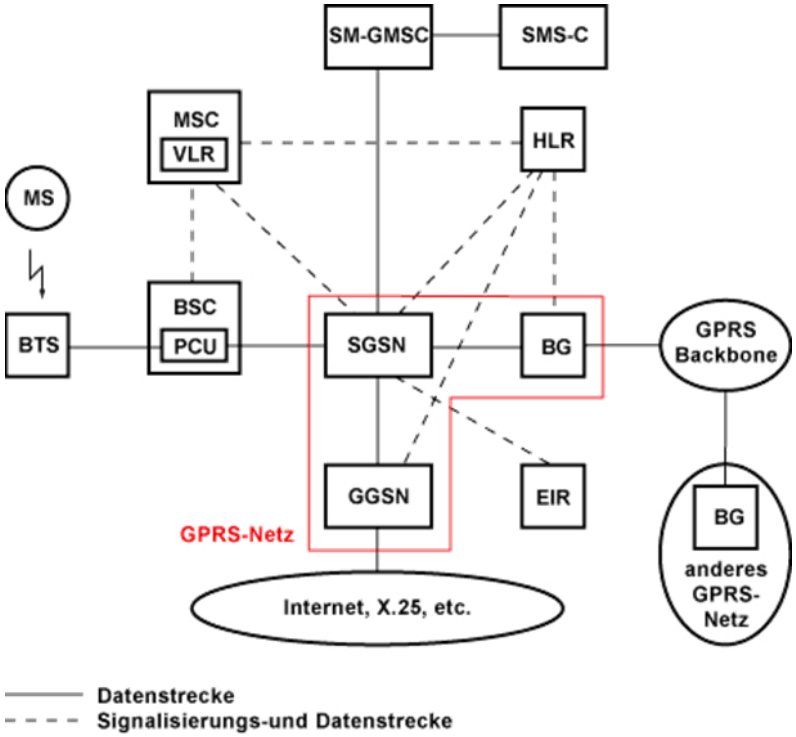


Abbildung 9.4: Die Architektur von GSM ergänzt mit den GPRS Komponenten [23]

### 9.4.1 GPRS Architektur

Um GPRS nutzen zu können, braucht der Benutzer, zusätzlich zu den Vorraussetzungen für GSM, nur ein GPRS-fähiges Endgerät. Der Provider muss allerdings einige Änderungen an seinem GSM System vornehmen. Die Hauptkomponenten sind die sogenannten GPRS-Support-Nodes (GSN). Im Wesentlichen gibt es zwei solcher Support-Nodes. Der GGSN (Gateway GPRS Support Node) etwa ist für die Anbindung an fremde Netze vorgesehen und fungiert zusammen mit dem Border Gateway als Schnittstelle zu externen Netzen. Vorraussetzung ist, dass das fremde Netz ebenfalls paketvermittelnd ist. Der SGSN (Serving GPRS Support Node) ist für die Mobilität des Mobiltelefons, die Verwaltung der Sessions, die Authentifizierung und die Paketvermittlung verantwortlich. Er übernimmt in etwa die Funktion des MSC bei Sprachanwendungen. Der SGSN ist auch für die Kommunikation mit dem Funknetz verantwortlich.

Von der mobilen Station (MS) werden die Daten an die nächste Basisstation (BTS) übertragen. Der Base Station Controller ist für die GPRS-Kanalvergabe an die Endgeräte verantwortlich. Die BTS und der BSC müssen für GPRS ebenfalls aufgerüstet werden. Wichtige zusätzliche Komponenten sind neue Channel-Kodierungen und Signalisierung. Im BSC befindet sich eine Paket Control Unit (PCU), die für die Datenpakete verantwortlich ist. Von dieser PCU gehen die Daten dann an den SGSN und via GGSN ins Internet oder via BG an einen anderen GPRS-Betreiber.

Dieser Überblick sollte ausreichen um die Sicherheitsaspekte für GPRS untersuchen zu können. Für weitere Informationen zur Architektur wende sich der interessierte Leser bitte an die entsprechende Fachliteratur. [24]

## 9.5 GPRS Sicherheit

Zuerst sollen nochmals die drei Hauptpunkte der Sicherheitsdiskussion in Erinnerung gerufen werden:

- Anonymität
- Vertraulichkeit
- Authentifizierung

Im Folgenden werden nun die von GPRS gebotenen Sicherheitsmechanismen anhand dieser Aspekte beschrieben. Anschliessend werden die verschiedenen sicherheitsrelevanten Interfaces und Komponenten von GPRS aus einer Angreifersicht untersucht. In einem dritten Schritt wird der Ende-zu-Ende Ansatz behandelt.

### 9.5.1 Sicherheitsmechanismen

#### Anonymität

Die Anonymität des Benutzers wird mit einem ähnlichen Ansatz versucht sicherzustellen, wie dies bereits aus dem GSM-Teil bekannt ist. Ziel ist es, einem Angreifer zu verunmöglichen, dass er über die Luftschnittstelle herausfinden kann, wer zurzeit gerade eine Datenverbindung unterhält. Als Identifikationsmerkmal für das GPRS-Netz dient wiederum die International Mobile Subscriber Identity (IMSI). Diese soll aber aus offensichtlichen Gründen nicht im Klartext über die Luftschnittstelle übertragen werden. GPRS benützt zu diesem Zweck die Temporary Logical Link Identity (TLLI). Diese wird der IMSI temporär zugeteilt und ist nur in einer Routing Area eindeutig. Einzig der SGSN und das Mobiltelefon selbst kennen die Beziehung zwischen IMSI und TLLI. Wenn eine neue TLLI zugewiesen wird, wird diese vom SGSN an das Mobiltelefon verschlüsselt übermittelt.

#### Vertraulichkeit

Für GPRS wurde der aus GSM bekannte A5-Algorithmus weiterentwickelt. Das Funktionsprinzip und der Schlüsselaustausch sind identisch mit der Verschlüsselung auf GSM Basis. Informationen dazu sind dem Kapitel über die GSM Verschlüsselung zu entnehmen. Im Folgenden wird lediglich auf Unterschiede und Implikationen für den Datenverkehr eingegangen. Der Algorithmus wird systemseitig nicht mehr im BSC angewandt, sondern erst im SGSN. Somit bietet GPRS eine längere verschlüsselte Strecke. Allerdings sollte man dies nicht überbewerten, denn erstens kann man nicht sagen, wo genau der SGSN steht und zweitens ist damit immer noch nicht das ganze System abgedeckt.

Der SGSN bekommt den Schlüssel Kc zusammen mit den Authentifizierungsdaten vom AuC und startet die Verschlüsselung nach dem „Authentification and Ciphering Request“.

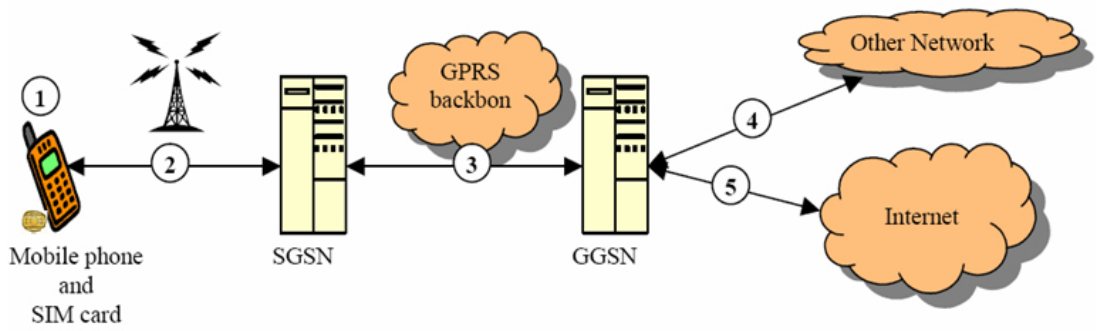


Abbildung 9.5: Die verschiedenen Angriffspunkte, die einem potentiellen Angreifer zur Verfügung stehen [21]

Die Mobilstation beginnt mit der Verschlüsselung sobald sie die Antwort auf den Request gesendet hat. Da der GPRS-A5 Algorithmus nicht öffentlich zugänglich ist, können an dieser Stelle keine genauen Aussagen zu den Sicherheitseigenschaften gemacht werden. Es ist jedoch davon auszugehen, dass die Sicherheit in etwa jener von GSM-A5 entspricht. Es soll aber ein längerer Schlüssel verwendet werden.

#### Authentifizierung

Auch die Authentifizierung wurde mit GPRS nicht auf den Kopf gestellt. Man stützte sich dabei auf den vorhandenen GSM-Authentifizierungsvorgang. Es wurde einzig ein neuer A3-Algorithmus für GPRS entwickelt, dessen Spezifikationen jedoch nicht öffentlich sind [3]. Der SGSN bekommt vom Home Location Register und vom Authentication Center ein Triplet mit einer Zufallszahl, einer Signed Response und einen Schlüssel für die Verschlüsselung. Mit den ersten beiden Komponenten wird wieder in einem Challenge-Response-Verfahren die Authentizität des Benutzers sichergestellt. Das Mobiltelefon rechnet mit dem persönlichen Schlüssel Ki und der Zufallszahl mittels des bekannten Algorithmus A3 die Signed Response und schickt sie dem Netz zurück. Einziger Unterschied zur GSM-Authentifizierung ist, dass der Authentifizierungsvorgang neu im SGSN abgewickelt wird, welcher nun für alle drei Sicherheitsaspekte von zentraler Bedeutung ist. [20]

### 9.5.2 Analyse anhand der Angriffspunkte

Um ansatzweise eine umfassende Analyse der Sicherheitsaspekte vornehmen zu können, ist es erforderlich die Perspektive eines Angreifers einzunehmen und so die verschiedenen Angriffspunkte zu untersuchen.

Für den Angreifer gibt es folgende Interfaces, die für einen Angriff in Frage kommen:

- Die Mobilstation und die Luftschnittstelle
- GPRS Backbone Netz

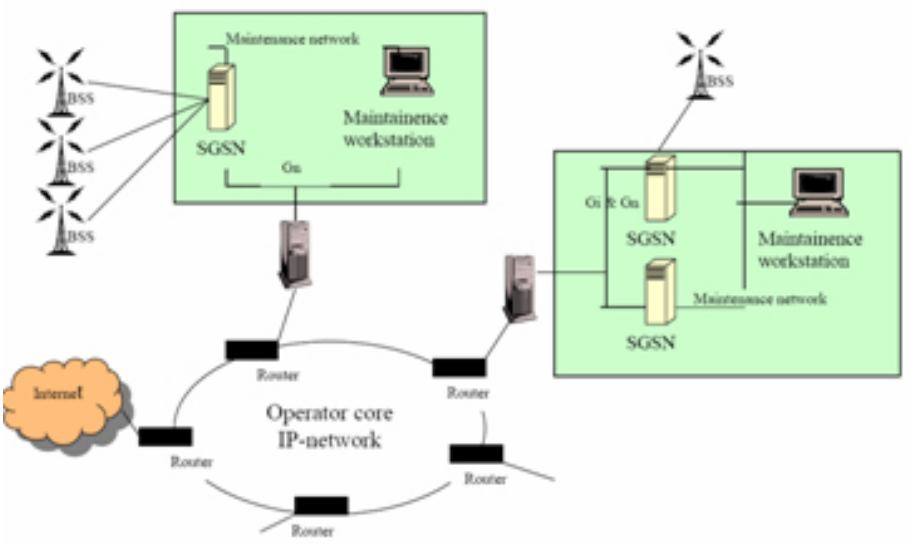


Abbildung 9.6: Ein Beispiel eines GPRS Backbones [21]

- Verbindung zu anderen Providern
- Verbindung in öffentliche Netze

Die Mobilstation und die Luftschnittstelle wurden im Wesentlichen bereits im GSM-Teil dieser Arbeit behandelt und werden hier nicht weiter ausgeführt. Die anderen Punkte werden nun der Reihe nach analysiert und es wird versucht Lösungsansätze aufzuzeigen.

### GPRS Backbone Netz

Aus Untersuchungen ist bekannt, dass ein beträchtlicher Teil aller Angriffe auf Informationsnetze aus dem Innern des Netzes erfolgt. Es ist deshalb nicht nur die Schnittstelle zwischen der MS und dem GPRS-Netz zu betrachten, sondern das gesamte Backbone Netz des Betreibers.

Für den Backbone-Bereich gibt es keine verbindlichen Richtlinien seitens des GPRS Standards. Grundsätzlich liegt die Sicherheit völlig im Ermessen des Netzbetreibers. Ziel der Sicherheitsüberlegungen an dieser Stelle ist es, sicherzustellen, dass niemand unauthorisiert Zugang zu den standardmässig unverschlüsselten Daten erhält. Dies betrifft also vor allem die Vertraulichkeit und die Anonymität.

Zwischen dem SGSN und dem GGSN kommt das GPRS Tunneling Protocol (GTP) zum Einsatz. GTP ist ein Protokoll auf IP-Basis, mit dem unterschiedlichste Protokolle transportiert werden können: DNS Anfragen, Zonenübergänge, HTTP oder FTP Verkehr und einige mehr. Standardmässig ist GTP aber nicht verschlüsselt. Da bekanntlich die A5 Datenverschlüsselung am SGSN wieder aufgehoben wird, fliesst auch der gesamte Datenverkehr zwischen SGSN und GGSN unverschlüsselt durch das Netz.

Zu beachten ist ausserdem, dass ein SGSN nicht nur zu einem GGSN eine Verbindung unterhält und umgekehrt, sondern dass beliebig viele SGSN mit beliebig vielen GGSN

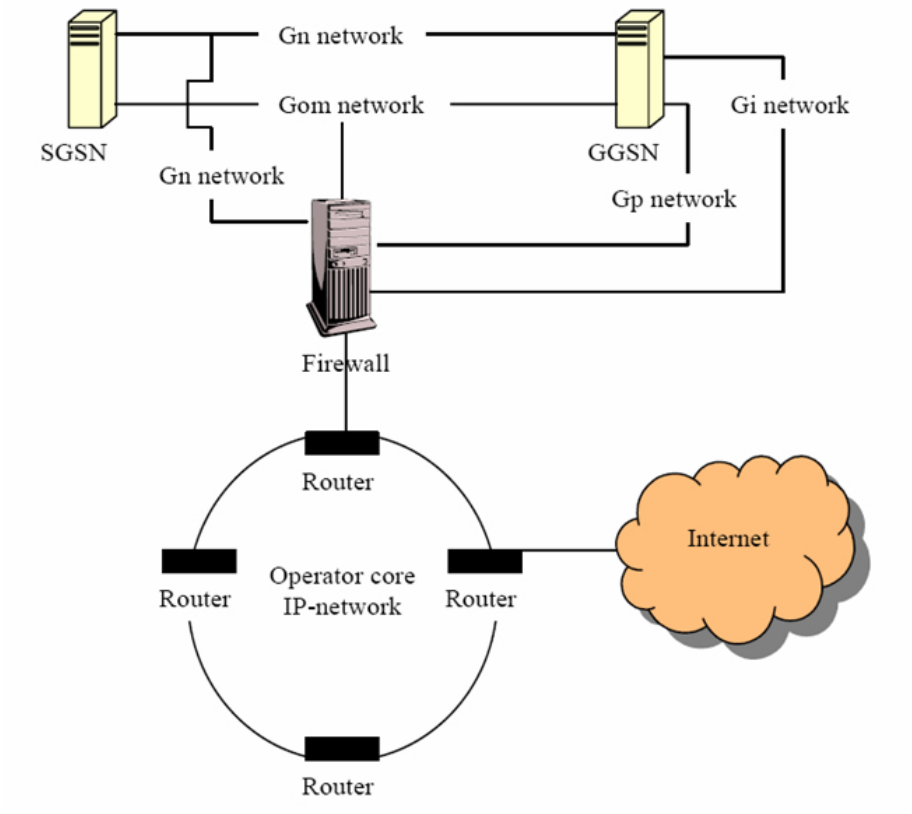


Abbildung 9.7: GPRS Backbone mit Firewall [21]

eine Verbindung unterhalten können. Dies vor allem im Falle von Roaming, welches im nächsten Abschnitt untersucht wird.

Da GPRS häufig ein Zusatz zu einem bestehenden GSM-Netz ist, wird für den GPRS Backbone häufig ein bestehendes IP-Netzwerk verwendet, über welches potentiell unzählige andere Anwendungen laufen. Dies vergrössert die Menge an potentiellen Angreifern, die Zugang zu diesem Netz haben und schafft so zusätzliche Eintrittspunkte für externe Angreifer.

Als Lösung könnte man sich hier Firewalls vorstellen, die wie in Abbildung 9.7 gezeigt, die wesentlichen Teile – vor allem den Zugang zu den MS und die Support-Nodes – schützen. Da es sich hier um nicht GPRS-spezifische Probleme handelt, sei der Leser an die entsprechende Fachliteratur zu Computernetzwerksicherheit verwiesen [22].

### Verbindung zu anderen Providern

Um Roaming anbieten zu können, muss eine Verbindung zwischen den Providern bestehen. Diese kann über eine spezielle Verbindung oder aber über jedes beliebige IP-basierte Netz geschehen. Hierbei ist natürlich vor allem das Internet gemeint. Auf Abbildung 9.8 ist ein Schema einer solchen Zusammenarbeit erkennbar. Der Benutzer befindet sich im unteren Netz als Gast. Der SGSN des besuchten Netztes baut nun eine Verbindung zum GGSN des Heimnetzes auf, von wo aus das Internet zugänglich ist.

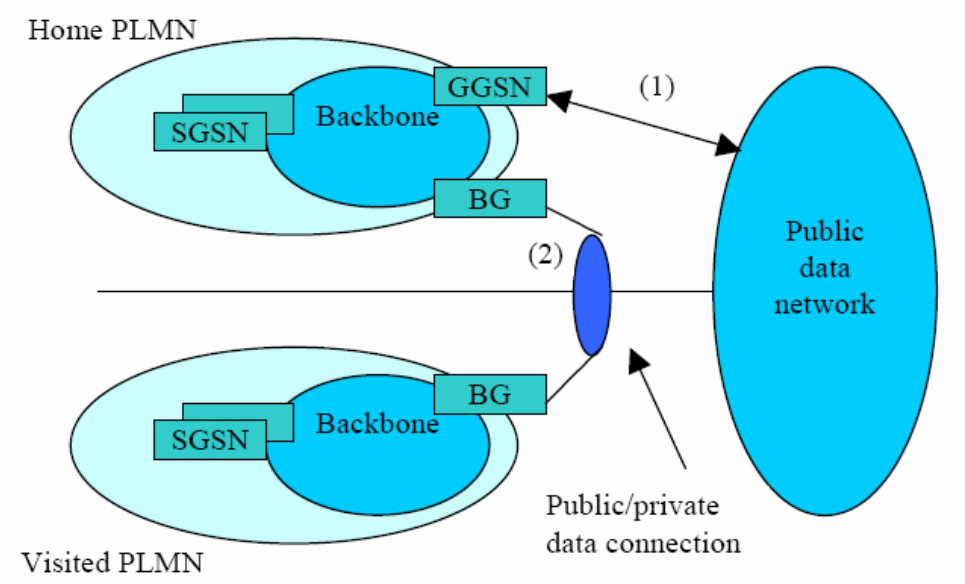


Abbildung 9.8: Das Schema einer Roamingverbindung [21]

Sicherheitsrisiken ergeben sich schon allein aus der Frage, ob dem Service-Provider gegenüber vertraut werden kann. Bei staatlicher Einmischung oder korrupten Wirtschaftssystemen nützen die besten vereinbarten technischen Schutzmechanismen nichts, wenn sich die Gegenpartei nicht daran hält oder die Mechanismen systematisch unterläuft und ausschaltet. Die Angriffsziele liegen hier in einem ähnlichen Bereich, wie beim GPRS-Backbone: Abhorchen der Daten, Analyse des Datenverkehrs und Manipulation der Daten zielen auf die Hauptpunkte Anonymität und Vertraulichkeit. Um diesen Angriffszielen entgegenzuwirken können die GPRS-Netzbetreiber bilateral sichere Verbindungen eingehen, zum Beispiel über IPSec (siehe weiter unten). Dies löst die Probleme, die auf der Strecke zwischen den Providern auftauchen, aber nicht die Frage nach der Vertrauenswürdigkeit des Gegenübers. Diese lässt sich nicht mit technischen Mitteln sicherstellen und unterliegt zutiefst der menschlichen Eigenart als das schwächste Glied in der Kette (siehe weiter unten). Bei IPSec ist ausserdem zu beachten, dass es sich um eine vorgeschlagene Option handelt, die von den Providern nicht zwingend verwendet wird [22].

### Verbindung mit öffentlichen Netzen

Der Nutzen von GPRS entsteht erst mit der Anbindung an das öffentliche Internet. Doch überall wo Systeme gegen aussen geöffnet werden, entsteht ein zusätzliches Sicherheitsrisiko. Gegen aussen wirkt der GGSN wie ein gewöhnlicher IP-Router und kann deshalb von aussen auch gesehen werden. Meist werden intern und extern andere IP-Adressen vergeben. Meldet sich ein Benutzer am GPRS-Netz an, bekommt er eine netzinterne IP-Adresse vom Provider zugewiesen. Möchte der Benutzer nun auf das Internet zugreifen, wird seine IP-Adresse am GGSN in jedem IP-Paket mittels NAT in eine externe, dem Provider zugewiesene, IP-Adresse geändert. Wenn nun ein Datenpaket zurückkommt geschieht dasselbe in umgekehrter Richtung. Zusätzlich übernimmt der GGSN eine Firewallfunktion gegen aussen.

Hauptziel des Angreifers von aussen ist vor allem dem Benutzer zu schaden, indem er ihm grosse Datenpakete schickt, welche wiederum eine hohe Rechnung verursachen, oder indem er versucht Informationen zu stehlen, welche ihm zum Reputationsaufbau unter Hackern dient, oder welche er verkaufen kann. Der Schutz dieses sensiblen Bereiches erfolgt ebenfalls nach den Prinzipien des Schutzes eines normalen Computernetzwerks, wie wir es zum Beispiel in Firmen finden. Deshalb sei an dieser Stelle wiederum an die entsprechende Fachliteratur verwiesen. Spezielle Sicherheitsmassnahmen im Ende-zu-Ende Bereich finden sich im nächsten Kapitel [22].

### 9.5.3 Ende-zu-Ende Sicherheit

Wenn man die wesentlichen Sicherheitsaspekte Authentifizierung, Vertraulichkeit und Integrität betrachtet, so musste weiter oben festgestellt werden, dass die GPRS-Architektur diesen Anforderungen im Ende-zu-Ende Bereich nicht gerecht werden kann.

Im Folgenden wird deshalb ein anderer Ansatz beschrieben, um diese Sicherheit zu erreichen. Bis jetzt wurde immer davon ausgegangen, dass der Netzbetreiber oder das Netz selbst für Sicherheit zu sorgen habe. Ein weiterer Ansatz, der in der Informationstechnologie viel diskutiert wird, ist die Sicherstellung der Sicherheit durch den Benutzer selbst und zwar durch die gesamte Informationskette (Ende-zu-Ende). Die zwei Technologien, die dazu in Frage kommen, sind Virtual Private Network (VPN) und IPSecurity (IPSec) [25]. Beide wurden ursprünglich für die verkabelte Welt entwickelt. Mit der zunehmenden Verschmelzung der verkabelten mit der kabellosen Welt und unter Berücksichtigung der Tatsache, dass sie heute zu den sichersten Protokollen für IP-Sicherheit zählen, werden sie auch für GPRS-Anwendungen interessant. Beide sollen im Folgenden kurz umrissen werden und anschliessen soll ihre Anwendbarkeit für GPRS diskutiert und analysiert werden.

#### Virtual Private Network

Virtual Private Networks (VPN) erlauben eine Verschlüsselung von Daten und eine sichere Übermittlung über öffentliche Netze, wie zum Beispiel das Internet. VPNs arbeiten auf einem Ende zu Ende Konzept. Das heisst der eine Endpunkt verschlüsselt die Daten, welche dann erst wieder am anderen Endgerät entschlüsselt werden. VPN-Verschlüsselungen werden auch als VPN-Tunnels bezeichnet. Dies weil meist das gesamte IP-Packet (mit Header) verschlüsselt wird und anschliessend in ein neues Packet geschrieben wird. Die zur Verschlüsselung verwendete Technologie ist meist IP Security, welche im folgenden Abschnitt beschrieben wird [26].

#### IP Security

IPSec setzt auf der Internetschicht des TCP/IP Protokolls an. Im Wesentlichen deckt IPSec Verschlüsselung und Integritätsschutz für IPv4 und IPv6 ab [25]. IPSec kennt zwei unterschiedliche Sicherheitsdienste: Zum einen das sogenannte Authentication Header (AH)

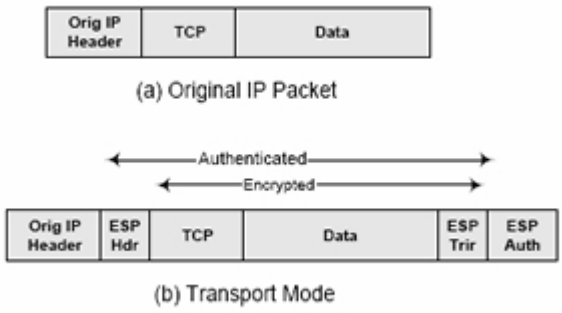


Abbildung 9.9: Die Funktionsweise von IPSec im Transportmodus [25]

und zum anderen Encapsulating Security Payload (ESP). AH bietet nur Ursprungsauthentifikation und Datenintegrität, aber keine eigentliche Verschlüsselung. Mit ESP stellt IPSec zusätzlich die Vertraulichkeit der Daten sicher. Da aber gerade der letzte Aspekt für die Sicherheitsdiskussion von enormer Bedeutung ist, beruhen die weiteren Ausführen, falls nicht anders erwähnt, auf dem ESP-Konzept.

Beide Arten von IPSec kennen zwei Betriebsarten: Im Transportmodus werden grundsätzlich die transportierten Daten geschützt (Payload) und im Tunnelmodus wird das gesamte IP-Paket geschützt und in einem neuen IP-Paket verpackt. In der Grafik 1.9 wird gezeigt, wie ein einzelnes IP-Paket mit IPSec ESP im Transportmodus verpackt wird. Das gesamte Paket – mit Ausnahme des IP-Headers – wird verschlüsselt. Dazu wird zur Authentizitätssicherung ein ESP Header und Trailer geschrieben. Das gesamte Paket wird dann in ein neues IP-Paket verpackt und verschickt. Diese kurze Beschreibung soll für unsere Zwecke genügen. Für mehr Informationen, im Speziellen zum Schlüsseltaustausch und den unterliegenden Verfahren, wende man sich an die Fachliteratur.

### VPN und IPSec in GPRS

Zuerst wird anhand eines exemplarischen Beispiel die Anwendung von VPN und IPSec aufgezeigt und anschliessend werden die Mechanismen im Zusammenspiel mit GPRS analysiert. Vor allem in Geschäftsbereichen wird der sichere mobile Datenaustausch, zum Beispiel mit einem Firmennetzwerk, immer wichtiger. Als Endpunkte der sicheren Verbindung sollen deshalb in unserem Beispiel ein Mobiltelefon mit Datenanwendungen und ein Firmennetzwerk dienen. Die beiden Komponenten können aber beliebig durch andere Geräte mit ähnlichen Charakteristiken ersetzt werden.

Um eine sichere Verbindung zum Firmennetzwerk aufzubauen, meldet sich der Benutzer ganz normal bei seinem GPRS-Netz an und baut somit auch eine logische Verbindung zum SGSN auf. Im Folgenden wird vom SGSN zum GGSN ein GTP Tunnel aufgebaut und dem Mobiltelefon eine IP-Adresse zugewiesen. Danach kann der Client zum Endpunkt (zum Beispiel einem Security Gateway oder Proxy) eine sichere Verbindung über VPN aufbauen. Bei IPv4 ist davon auszugehen, dass sowohl der GGSN als auch der Security Gateway Network Address Translation (NAT) verwenden um interne und externe IP-Adressen zusammenzuführen. Es muss dabei beachtet werden, dass IPSec praktisch nur

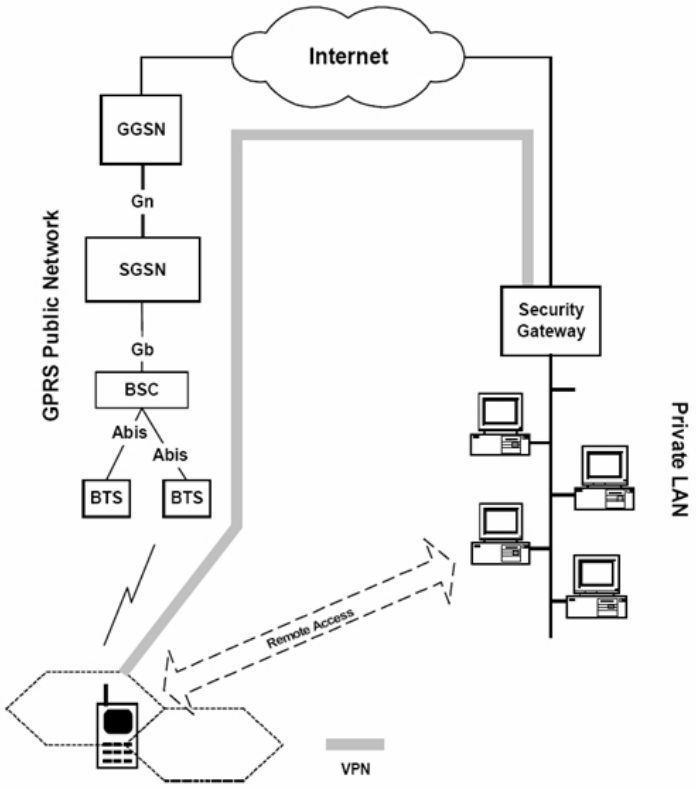


Abbildung 9.10: Das Schema einer VPN-Verbindung in ein Firmennetzwerk [25]

im Transportmodus arbeiten kann, da der Tunnelmodus von den Mobiltelefonen praktisch nicht unterstützt wird. Ausserdem wird mit dem Tunnelmodus das Datenvolumen erhöht, was zusätzlich negative Implikationen auf die Übertragungsgeschwindigkeit und die Kosten hat.

Ein Datenpaket aus unserem Beispiel wird nun zuerst durch VPN und IPSec und danach durch die normalen GPRS-Sicherheitsmechanismen verschlüsselt. Es gelangt so doppelt verschüsselt von dem Mobiltelefon bis zum SGSN. Der SGSN hebt nun die GPRS-Verschlüsselung wieder auf und füllt die Datenpakete ins GPRS Transport Protocol (GTP) ab, welches bekanntlich keine zusätzliche Verschlüsselung kennt. Mittels GTP werden die Daten dann vom SGSN zum richtigen GGSN gesendet. Dieser holt aus dem GTP wiederum die Datenpakete heraus und wendet eine Network Address Translation (NAT) an, um die interne IP-Adresse in die dem Provider zugewiesene externe IP-Adresse umzuwandeln. Übers öffentliche Netz erreichen die verschlüsselten Datenpakete nun den Security Gateway des Firmennetzwerks, der die IPSec-Verschlüsselung wieder rückgängig machen und das Paket dem richtigen Teilnehmer zuweisen kann. In der umgekehrten Richtung funktioniert der Mechanismus ganz ähnlich. Wenn der GGSN ein verschlüsseltes Paket erhält, muss ebenfalls NAT angewendet werden um das Paket intern (oder auch extern im Falle von Roaming) dem richtigen SGSN zuzustellen. Die Luftübertragung funktioniert dann wieder gleich wie im ersten Fall.

Mit dieser Technologie sind aber auch verschiedene technische Schwierigkeiten verbunden. Zwei davon, sind von zentraler Bedeutung. Einerseits schlägt das GTP-Tunnelprotokoll

bei VPN mit einer schlechten Performance und einem hohen Ressourcenaufwand zu Buche und anderseits bringt NAT in Zusammenarbeit mit VPN einige Probleme mit sich. Insbesondere kann es Inkompatibilitäten zwischen IPSec und NAT geben. Zur Vertiefung dieses Themas und den Lösungen sei das Paper von Xenakis [25] empfohlen.

Der Hauptvorteil von VPN ist ohne Zweifel die zur Zeit bestmögliche Sicherheit im Ende zu Ende Bereich. Ausserdem kann VPN mit geringen Anpassungen seitens des GPRS-Netzbetreibers eingesetzt werden, was bei anderen Sicherheitstechnologien nicht unbedingt der Fall ist. Mit VPN entstehen dem Benutzer auch nicht mehr Kosten. Da alle Einrichtungen vom Benutzer und dem Firmennetzwerk zur Verfügung gestellt werden müssen und nicht mehr Verkehr über das GPRS-Netz geht, kann der Netzbetreiber auch nicht mehr Gebühren kassieren.

Die beschriebene Technologie bringt aber auch einige Nachteile mit sich. So obliegt die Sicherstellung der Übertragungssicherheit plötzlich beim Benutzer, der nicht in jedem Fall die nötige Qualifikation oder Motivation mitbringt, um sich den Sicherheitsbedürfnissen bewusst zu sein. Es muss ausserdem auf jedem Mobiltelefon eine spezielle VPN-Software installiert werden, was wiederum eine zusätzliche Eintrittshürde darstellt. Im Weiteren sei hier noch auf die Charakteristiken von mobilen Endgeräten hingewiesen. Es steht eine geringe Prozessorleistung zur Verfügung, die Batterie hat eine beschränkte Laufzeit, Speicherkapazität ist noch teuer und die Latenz in GPRS-Netzen ist noch relativ hoch. All diese Faktoren haben Einfluss auf die VPN-Anwendung auf Mobiltelefonen. VPN ist rechenintensiv und braucht deshalb auch mehr Energie und Prozessorleistung, IPSec braucht relativ viel Speicherplatz und die VPN-Anwendung erhöht die Latenz im GPRS-Netz noch zusätzlich. [25]

## 9.6 Allgemeine Sicherheitsfragestellungen in Mobilfunknetzen

Dieser Abschnitt behandelt zwei Probleme, die im Gebiet der Sicherheit von Mobilkommunikation eine Rolle spielen. Sie sind jedoch nicht eindeutig GSM oder GPRS zuordenbar, sondern betreffen beide Gebiete. Deshalb werden das Datenschutzproblem und das „Problem Mensch“ an dieser Stelle einer gesonderten Betrachtung unterzogen.

### 9.6.1 Datenschutzproblem

Der Datenschutz ist ein Problem, das viele Bereiche des heutigen, modernen Lebens umfasst. Nämlich alle Bereiche in denen Kundendaten anfallen, welche in einer Datenbank erfasst werden und die durch ausgeklügelte Data Mining- oder Data Warehouse Methoden ausgewertet werden können. Solche Daten fallen nun auch bei den Betreibern von Mobilfunknetzen an und eine ganze Reihe von entsprechenden Auswertungsmöglichkeiten ist denkbar. Personalisierte Kundenprofile können erstellt werden oder die Aufenthaltsorte der Personen sind ermittel- und verfolgbar. Heutzutage sind solche Vorgehensweisen in der

Regel von Gesetzes wegen verboten. Doch an eine strikte Einhaltung der Gesetze von jedermann kann nicht geglaubt werden. Dies führt zu der Annahme, dass dort, wo sensitive Daten anfallen, diese auch missbraucht werden können. Im Gegensatz zu anderen Orten an denen sensitive Daten anfallen (z.B. der Einkauf) hat der Kunde bei der Mobiltelefonie keine Möglichkeit seine Daten zu verweigern und genauso wenig ist er über potentielle Auswertungen informiert. Dieser Zustand klingt relativ beunruhigend und ist eigentlich eine näheren Betrachtung wert, welche jedoch den Rahmen dieser Seminararbeit zu sehr dehnen würde.

Weiter sei an dieser Stelle auch darauf hingewiesen, dass sowohl GSM, als auch GPRS über spezielle Lawful-interception-Interfaces verfügen, um das offizielle Abhorchen und Auswerten durch Staatsorgane zu ermöglichen. Diese absichtliche Sicherheitslücke kann in der Verbrechensbekämpfung ein Segen sein, kann aber den Datenschutz massiv untergraben. So ist nicht sichergestellt, dass diese Schnittstelle nicht für andere Zwecke missbraucht wird, oder Behörden präventiv Daten sammeln. Dies mag nach Verschwörungstheorie tönen, kann besonders in totalitären Staaten durchaus zum Tagesgeschäft gehören. Weitere Ausführungen wären Spekulation. [27]

### 9.6.2 Problem Mensch

Eine weitere Facette der vielseitigen Sicherheitsdiskussion liegt in der Betrachtung des einzelnen Benutzers eines mobilen Telekommunikationssystems. ”Jede Kette ist so stark wie ihr schwächstes Glied”, sagt ein berühmtes Zitat und in vielen Bereichen ist der Mensch selbst dieses schwächste Glied. Im Umfeld der Mobiltelefonie kann das an einfachen Beispielen aufgezeigt werden: Leute deaktivieren ihren PIN im Telefon aus Bequemlichkeit oder sie führen eigentlich vertrauliche oder private Gespräche in aller Öffentlichkeit in einer Lautstärke, dass nichts davon vertraulich oder privat bleibt.

Somit werden angebotene Sicherheitsvorkehrungen einfach nutzlos gemacht, meistens sogar völlig unwissentlich. Dies ist vor allem auf die mangelnde Sicherheitssensibilität vieler Benutzer zurückzuführen. Aus diesem Grunde sei erwähnt, dass die besten Absichten bezüglich Sicherheitsvorkehrungen nichts nützen, wenn sie einfach ignoriert werden. Aus diesem Grunde wäre eine bessere Information und Aufklärung der Benutzer ein guter und einfacher Weg, um mit relativ geringem Aufwand einen besseren allgemeinen Sicherheitslevel anbieten zu können.

## 9.7 Schlussbetrachtungen

Das Ziel dieser Arbeit war es, die GSM und GPRS zu Grunde liegenden Sicherheitsmechanismen zu ergründen und diese aufzuzeigen. Danach folgte eine eingehende Beschreibung von netzspezifischen, wie auch von mobilfunk-allgemeinen Problemen und Sicherheitslücken und, wo vorhanden, die Vorstellung möglicher Lösungskonzepte.

Dabei wurden zum Teil massive Lücken und Sicherheitsmängel aufgezeigt, welche die Sicherheit und das breite Vertrauen der über einer Milliarde Kunden in das System in Frage

stellen. Zwar wurde gezeigt, dass bei den Spezifikationen ein Augenmerk auf die Sicherheit geworfen wurde und demzufolge Sicherheitskonzepte vorhanden sind, doch einige Punkte wurden schlicht zu wenig beachtet oder einfach nur schlecht umgesetzt in den laufenden Netzen. Ein gewichtiger Punkt der an dieser Stelle erwähnt sein muss, ist, wie wenig überhaupt über diese teilweise recht eklatanten Schwächen in der Öffentlichkeit bekannt ist. Über eine Milliarde Kunden vertrauen den Systemen mehr oder weniger blind, ohne einen Gedanken an die Sicherheit der geführten Gespräche oder der gesendeten Daten zu verschwenden. Dies führt gerade zur Fragestellung, ob die breite Masse überhaupt ein hohes Mass an Sicherheit wünscht oder ob sie mit den momentanen Lücken leben kann. Die gesamte Problematik wirft auch ein schiefes Licht auf die Netzbetreiber und auf die Hardwarehersteller, welche sich zu wenig engagiert zeigen in der Behebung offenkundiger Mängel.

Natürlich ist an dieser Stelle wieder einmal eine alte Gleichung abzurufen: Sicherheitsmassnahmen benötigen Rechenressourcen und verursachen Kosten. Dieser Tradeoff zu machen und am Ende alle beteiligten Parteien zufrieden zu stellen erfordert viel Geschick. Und wer weiss, vielleicht ist der momentane Zustand der ideale Sicherheitseinsatz für ein weltweites Kommunikationssystem? Wie auch immer, dieser Seminarbericht zeigt auf, welche Stellen kritische Lücken bestehen. Ob die Benutzer diesen Zustand viel länger hinnehmen, oder sogar immer mit den Problemen leben können, ist eine andere Frage.

Wir möchten uns bei Prof. Dr. Burkhard Stiller für die Betreuung und Hilfsbereitschaft und anschliessende Korrektur herzlich bedanken.

# Literaturverzeichnis


[1] G. Heine: GSM Networks: Protocols, Terminology, and Implementation;
Artech House Publishers, Boston, Massachusetts, U.S.A., 1999.

[2] C. Brookson: GSM (and PCN) Security and Encryption;
http://www.brookson.com/gsm/gsmdoc.pdf, 1994.

[3] J. Quirke: Security in the GSM system, 2004;
http://www.ausmobile.com/downloads/technical/Security in the GSM system 01052004.pdf.

[4] J. Schiller: Mobilkommunikation 2., überarbeitete Auflage;
Pearson Studium, Deutschland, 2003.

[5] GSM-World- Webseite;
http://www.gsmworld.com, besucht Mai 2005.

[6] The Clone: The Security Technical Whitepaper for 2002;
http://www.hackcanada.com/blackcrawl/cell/gsm/gsm security.html, 2002.

[7] D. Margrave: GSM Security and Encryption;
http://www.hackcanada.com/blackcrawl/cell/gsm/gsm-secur/gsm-secur.html

[8] Chaos Computer Club e.V.: GSM: Security by obscurity, Datenschleuder Nr 63;
http://www.chscene.ch/ccc/ds/63/020 gsm.html

[9] M. Briceno, I.Goldberg, D.Wagner: An implementation of the GSM A3A8 algorithm (Specifically, COMP128);
http://www.gsm-security.net/papers/a3a8.shtml. 1998.

[10] Wikipedia: A5 (Algorithmus);
http://de.wikipedia.org/wiki/A5 Algorithmus, besucht Mai 2005.

[11] Alex Biryukov, Adi Shamir, David Wagner:
Real Time Cryptanalysis of A5/1 on a PC;
http://cryptome.org/a51-bsw.htm, April 2000.

[12] 3rd Generation Partnership Project:
3GPP TS 55.216 V6.2.0 (2003-09) Technical Specification
http://www.gsmworld.com/using/algorithms/docs/a5_3_and_gea3_specifications.pdf, 2003.

[13] Infoserver IT-Security: GSM-Sicherheit, Lexikoneintrag;
http://www.infoserversecurity.org/itsec infoserver v0.5/sections/home/
sections/links/1074011155/index.html, besucht Mai 2005.

[14] Isaac: GSM Cloning;
http://www.isaac.cs.berkeley.edu/isaac/gsm.html; besucht Mai 2005.

[15] GSM secutity.net: FAQ: "Have the A3 and A8 algorithms been broken?";
http://www.gsm-security.net/faq/gsm-a3-a8-comp128-broken-security.shtml;
besucht Mai 2005.

[16] N. Katugampala, S. Villette, A. M. Kondoz:
Secure Voice over GSM and Other Low Bit Rate Systems;
http://www.ee.surrey.ac.uk/Personal/N.Katugampala/pubs/iee03.pdf.

[17] Sectra: NSK 200;
http://www.ee.surrey.ac.uk/Personal/N.Katugampala/pubs/iee03.pdf;
besucht Juni 2005.

[18] Orbit-IEX Webseite;
http://www.orbit-iex.ch, besucht April 2005.

[19] Geschäftsbericht Swisscom Mobile, 2003;
http://www.swisscom.com/SCMCMS/GB/gb03/, besucht April 2005.

[20] C. Brookson: GPRS Security, 2001;
http://www.brookson.com/gsm/gprs.pdf, besucht April 2005.

[21] http://student.grm.hia.no/master/ikt01/ikt6400/ekaasin/Master Thesis Web.pdf,
besucht April 2005.

[22] Alan Bavosa, Juniper Networks:
White Paper GPRS Security Threats and Solution Recommendations, 2004

[23] Elektronikkompendium: GPRS;
http://www.elektronik-kompendium.de, besucht April 2005.

[24] GSM World: What is GPRS?;
http://www.gsmworld.com/technology/gprs/intro.shtml, besucht Juni 2005.

[25] Christos Xenakis, Evangelos Gazis and Lazaros Merakos:
Secure VPN Deployment in GPRS Mobile Networks, 2003.

[26] Infoweek: Sicherheit der Kommunikationsnetze;
http://www.infoweek.ch, besucht April 2005.

[27] GI Gesellschaft für Informatik e.V: Wireless Internet Security;
http://www.gi-ev.de/informatik/lexikon/inf-lex-wireless-int-sec.shtml, besucht Juni
2005.

# Kapitel 10

# Diameter in Wireless Environments

*Christian Jaldon, Lukas Schweizer*

*Mit der immer stärker werdenden Tendenz jederzeit und überall auf möglichst alle Informationen/Daten zugreifen zu wollen braucht es neben mobilen Endgeräten auch Systeme und Protokolle, die diesem Bedürfnis überhaupt gerecht werden können. Dabei stellt sich auch die Frage, was zu tun wäre in dem Fall wo am derzeitigen Standort das eigene Netz nicht verfügbar ist. Gerade wenn eine Netzinfrastruktur verfügbar wäre könnte diese doch benutzt werden.*
*Auf den folgenden Seiten soll* Diameter *als ein AAA-Protokoll (Authentifizierung, Autorisierung, Abrechnung) vor allem hinsichtlich dem mobilen Einsatz näher gebracht werden. Dieses Protokoll soll die Nachfolge von* RADIUS *antretten, welches zur Zeit die oben genannte Problematik zu lösen versucht, heute allerdings nicht mehr sämtlichen Anforderungen gerecht wird. Bei unseren Erläuterungen werden zuerst die Probleme und deren Lösung die im Zusammenhang mit AAA auftreten genauer betrachtet um danach deren Realisierung durch* Diameter *im mobilen Einsatz besser verstehen zu können. Anschliessend folgt eine Beurteilung hinsichtlich Flexibilität, Mobilität und Sicherheitsaspekten von* Diameter *bevor wir unsere Erläuterungen mit der Diameterimplementierung von* Mobile IP *schliessen.*

## Inhaltsverzeichnis

## 10.1 Einleitung

Die Nachfrage nach mobilen Diensten, ist enorm und stellt heute ein milliardenschwerer Markt dar. Die fortschreitende Verbreitung von mobilen Endgeräten, die anhaltende Tendenz nach permanenter Konnektivität und die steigende Komplexität der Netzwerke stellen hohe Ansprüche an die heutigen Kommunikationssysteme.

Die Kunden erhalten durch einen Internet Service Provider den Zugang zu solchen Diensten, wie Internet, E-Mail, Sprachdienste usw. Der Zugang erfolgt über das Heimnetz, der „Home Domain“ und ist meist an einem fixen Standort gebunden (Cable, ADSL). Durch die mobilen Endgeräte, Notebooks, PDA's, Smartphones etc., wollen die Kunden sich von der Abhängigkeit des Aufenthaltsortes loslösen und auf Dienste fremder Netzbetreiber zugreifen. Um aber eine möglichst flächendeckende Netzabdeckung anbieten zu können, müssen die verschiedenen Betreiber die Ressourcen partnerschaftlich nutzen und dabei aber die Kontrolle über den Zugang, die Nutzung und die Sicherheit behalten. Einerseits dürfen nur registrierte Benutzer vom Dienst profitieren. Andererseits muss stets Buchhaltung geführt werden über die Dauer der Nutzung und das Transfervolumen der Daten.

Die heutige Version von Mobile IP stellt keine Konzepte für Zugangskontrolle oder Verrechnung zur Verfügung. Neue Protokolle sollen die Anforderungen an eine Infrastruktur zur Authentifizierung, Authorisierung und Rechnungsstellung erfüllen. AAA-Protokolle (Authentication, Authorization and Accounting) werden heute eingesetzt, um die Problematik anzugehen.

RADIUS (Remote Authentication Dial In User Service), wurde 1992 entwickelt, um die AAA-Funktionalitäten verfügbar zu machen bzw. Network Access Servers um AAA-Funktionalitäten zu erweitern. Es ist heute noch weit verbreitet und findet vor allem bei ISP's für den Zugang zu den Hotspots Verwendung. Es funktioniert mit einer Client-Server-Architektur, wobei der RADIUS-Client für das Weiterleiten der Nachrichten an den RADIUS-Server verantwortlich ist. Dieser überprüft in einer Datenbank die Information auf ihre Richtigkeit und gewährt oder verbietet dem Client den Zugang.

RADIUS stosst in den heutigen Kommunikatinossystemen an seine Grenzen. Bedürfnisse wie Skalierbarkeit, Flexibilität oder Transportsicherheit können nicht mehr zufrieden stellend gedeckt werden. Gewisse Eigenschaften, wie der Datentransport per UDP, oder die mangelnde Fehlererkennung stellen sich als Schwächen heraus.

Diameter soll Abhilfe verschaffen. 1996 nahmen bereits namhafte Unternehmungen wie Cisco und Sun Microsystems an der Entwicklung für ein neues, besseres Protokoll teil. Ziel war es, das gegenwärtige RADIUS Protokoll so zu erweitern und zu Verbessern, dass vor allem auf die Schwächen reagiert werden konnte.

Diameter besteht aus dem Diameter Base Protocol und den Applications, den Erweiterungen für spezifische Anwendungen, wie Mobile IP, oder Diameter Network Access Server NASREQ. Im Gegensatz zu RADIUS verfolgt Diameter einen Peer-to-Peer-Ansatz. Somit ist es möglich, real-time Abfragen vom Server an die Clients zu machen. Insbesondere für Accounting-Funktionälitäten ist diese Architektur von grossem Vorteil.

## 10.2 Anforderungen an AAA-Systeme

AAA-Systeme – oder auch Triple-A-Systeme – werden beispielsweise von Internetdienstanbieter (ISP) verwendet, um sicherzustellen dass Personen, die einen Netzzugang wünschen identifiziert werden, ihnen entsprechend ihrer Identität Zugriff gewährt wird und Massnahmen getroffen werden welche dem ISP erlauben die erbrachte Leistung dem Kunden verrechnen zu können.

In diesem Abschnitt sollen die einzelnen Komponenten eines AAA-Systems erläutert werden um das Basisverständnis eines AAA-Prozesses für die später folgenden Ausführungen zur Implementierung von *Diameter* – als ein AAA-Protokoll für den mobilen Bereich – vermitteln zu können.

### 10.2.1 Aufgaben der einzelnen AAA-Komponenten

Bevor wir uns den konkreten Anforderungen an *Diameter* zuwenden können, sollen die jeweiligen Komponenten eines AAA-Systems erläutert werden.

#### Authentifizierung

Die *Authentisierung* ist das Nachweisen einer Identität, die *Authentifizierung* deren Überprüfung. Es gibt fünf verschiedene Ansätze diese Überprüfung durchzuführen. Der Nachweis kann erfolgen indem die Person zeigt, dass sie:

1. etwas habt (z.B. Hausschlüssel)
2. etwas weiss (z.B. Passwort)
3. etwas ist (z.B. Biometrische Merkmale)
4. an einem Ort ist (z.B. Rechneradresse)
5. etwas kann (z.B. Unterschrift)

Anhand einem oder auch durch Kombination mehrerer dieser Punkte kann eine Person identifiziert bzw. deren Identität nachgewiesen oder eben „authentisiert“ werden. Im Kontext der IT ist vor allem der Ansatz mit dem Passwort (eine Information, die nur der Benutzer selber haben sollte) und die Authentisierung anhand der Rechneradresse (z.B. Zugriffsberechtigung aller Arbeitsstationen aus einem bestimmten Subnetz auf den Druckerserver dieser Abteilung) verbreitet.

### Autorisierung

In unserem Kontext bezieht sich *Autorisierung* auf das Vergeben und Überprüfen von Zugriffsrechten auf Daten oder Dienste. Dieser Vorgang ist von Natur aus sehr stark mit der Authentifizierung gekoppelt. Bevor Rechte vergeben werden, ist es in der Regel äusserst vorteilhaft, wenn sichergestellt wird, dass auch die richtige Person diese Rechte erhält. Diese Zugriffsrechte stützen sich also auf eine – vorzugsweise nachgewiesene – Identität. Für sie wird vor einem Zugriff geprüft, ob und welcher Zugriff (z.B. lesend/schreibend) durchgeführt werden darf. Folgende Zugiffskontrollmodelle wurden in der Vergangenheit entwickelt:

1. Diskrete Zugriffskontrolle:

   Jeder Identität werden explizit Zugriffsrechte für Objekte erteilt.

2. Regelbasierte Zugriffskontrolle:

   Objekte werden eingestuft. Identitäten werden für gewisse Objekte bzw. Objektgruppen entsprechend deren Einstufung Zugriff erteilt oder verwehrt.

3. Rollenbasierte Zugriffskontrolle:

   Im Unterschied zur *diskreten Zugriffskontrolle* werden Zugriffsrechte nicht für alle Identitäten sondern für eine limitierte Anzahl Gruppen von Identitäten vergeben. Eine Identität wird ensprechend ihrer **Rolle**, welche sie im System einnimmt einer Gruppe zugewiesen.

Heute werden vorwiegend die *rollenbasierte Zugriffskontrolle* eingesetzt. Während bei der *diskrete Zugriffskontrolle* die Wartung und unter Umständen auch die Skalierbarkeit nicht ganz unproblematisch ist, kann der Ansatz der mit der *regelbasierten Zugriffskontrolle* verfolgt wird in der Praxis bestenfalls sehr eingeschränkt umgesetzt werden. Die Autorisierung auf Basis der *rollenbasierte Zugriffskontrolle* kann bei diesem flexiblen Ansatz mit relativ geringem Aufwand betrieben werden.

### Abrechnung

Unter Accounting oder unter dem deutschen Begriff *Abrechung*[1] fassen wir sämtliche Bestrebungen zusammen, die eine Verrechnung der tatsächlich verwendenten Ressourcen/Dienste ermöglicht. Das Hauptziel ist es die tatsächlich gebotene Leistung **verlässlich** verrechnen zu können. Dazu gehört beispielsweise das Festhalten von Startzeitpunkt eines Dienstes, dessen Ende sowie allfällige Unterbrüche und Fehlern (zum Beispiel soll im Falle eines Verbindungsabbruchs dieser entdeckt und die Verrechnung des Dienstes unterbrochen/gestoppt werden). Um die hohe Bedeutung der Verlässlichkeit zu verdeutlichen folgen hier zwei Beispiele:

[1]Das Wort „Verrechnen“ wäre ein wenig passender. Um aber nicht zusätzlich mit deutschen und englischen Begriffen die allgemeine Verwirrung zu erhöhen und um *AAA* auch in der deutschen Sprache als Abkürzung verwenden zu können wird hier das Wort „Abrechnen“ verwendet

Wird bspw. die physische Verbindung während einer Sitzung unerwartet getrennt, dann soll die Verrechnung des Dienstes nicht erst dann beendet werden, wenn der Benutzer sich bei der nächsten regulären Sitzung abmeldet.

Umgekehrt soll aber auch verhindert werden, dass die bewusste Trennung der Verbindung dazu führt, dass sämtliche in dieser Sitzung angefallenen Kosten nicht verrechnet werden, da sie nicht ordnungsgemäss beendet wurde.

Die gesammelten Informationen über die Benutzung kann auch der Kapazitätsplanung oder für eine genauere Berechnung der Kostenträger verwendet werden. Die Abrechnung selbst erfolgt in Echtzeit oder im Batch-Modus:

1. Echtzeit: Sämtliche abrechnungsrelevante Operationen führen sofort zu einer Übertragung ans System, welches die Verrechung übernimmt (idealerweise wird die Operation erst dann „freigegeben“ wenn die Bestätigung der Meldung eintrifft)
2. Batch: Informationen werden über eine definierte Zeitdauer hinweg gesammelt und erst bei Ablauf dieser Dauer übertragen

*Diameter* verwendet standardmässig den Echtzeit-Modus.

### 10.2.2 AAA-Anforderungen an *Diameter*

Es würde den Aufgabenbereich unserer Arbeit sprengen sämtliche Anforderungen an *Diameter* nieder zu schreiben. Stattdessen wollen wir einzelne Punkte nach [1] herausheben, welche sich als nötig oder äusserst nützlich für ein AAA-System erwiesen haben. Selbstverständlich müssen auch Massnahmen getroffen werden, um die oben unter Kapitel 10.2.1 genannte Funktionalität gewährleisten zu können.

#### Allgemeine Anforderungen

Neben den speziell im Fokus von AAA stehenden Anforderungen muss *Diameter* auch andere, allgemeinere Forderungen an ein Protokoll erfüllen, um überhaupt sinnvoll eingesetzt werden zu können. Da dies nicht der eigentliche Kernpunkt unserer Arbeit ist sollen auch hier nur die wichtigsten Forderungen kurz Erwähnung finden:

1. Skalierbarkeit: Die Forderung nach Skalierbarkeit ist insbesondere in einem extrem schnell wachsenden Netz wie dem Internet als äusserst wichtig einzuordnen. *RADIUS* zeigte diesbezüglich einige Schwächen. Durch folgende Punkte versucht *Diameter* dieser Forderung besser gerecht werden zu können:
   (a) Unterstützung von bis zu $2^{32}$ (ca. 4,29 Mrd) hängigen Anfragen
   (b) keine Benötigung von Ende-zu-Ende-Quittungen auf Ebene der Applikation (Bspw. von Diameter-Server $A$ zu Diameter-Server $B$) selbst.

   (c) Unterstützung von „Redirect Server“ (Umleiten auf anderen Server, z.B. für Load Balancing[2])

2. Robustheit: Fällt die Verbindung zu einem Endpunkt aus, **müssen** alle unquittierte Meldungen an einen alternativen Server gesandt werden (Failover[3]).

3. Erweiterbar/Flexibilität: *Diameter* soll bewusst mit der Idee einfach dienstespezifische Erweiterungen hinzufügen zu können (z.B. Mobile IP) realisiert werden.

4. Abwärtskompatibilität zu *RADIUS*: Bedauerlicherweise haben *Diameter* und *RADIUS* keine gemeinsame *PDU* (=Protocol Data Unit). Damit aber dennoch beide im gleichen Netz gemeinsam verwendet werden können, muss die Abwärtskompatibilität von *Diameter* zu *RADIUS* sichergestellt werden. Dies ist umso wichtiger da viele AAA-Bedürfnisse oftmals sehr spezifisch sind und deshalb Erweiterungen für bestehenden Anwendungen (unter *RADIUS*) implementiert wurden. In solchen Fällen wäre die Barriere für *Diameter* relativ hoch, was durch die Kompatibilität zu *RADIUS* und derer weiteren Verwendbarkeit umgangen werden kann.

5. Sicherheit:

   (a) Vertraulichkeit/Integrität der Datenobjekte: asymmetrische Verschlüsselung der Objekte durch eine Erweiterung [6] zu *Diameter*

   (b) Ermöglichen von Zertifikaten

6. Zuverlässigkeit: Accountinginformationen dürfen **nicht** verloren gehen. Die sichere Übertragung soll über die Protokolle SCTP (Stream Control Transmission Protocol) [7] oder TCP (Transmission Control Protocol) [8] geschehen.

7. Nutzbar unter IPv4 und IPv6

8. Unterstützung von Proxies und Broker

**Authentifizierung**

1. Re-Authentisierung on demand: Nach einer im Vorfeld definierten Zeitdauer verliert die getätigte Authentisierung ihre Gültigkeit und muss somit erneuert werden. Dies darf auch schon vor dem Ablauf des gewählten Intervalls geschehen. Dies ist z.B. dann hilfreich, falls keine ordnungsgemässe Abmeldung des Clients stattfand. Spätestens nach Ablauf des Zeitintervalls wird die Gültigkeit der Authentisierung aufgehoben, da auf Grund des Verbindungsabbruch keine Re-Authentisierung erfolgreich durchgeführt wurde. Oder anders herum: Auf Grund der Antwort auf eine Re-Authentisierung kann sich der Server vergewissern, dass der Client immer noch den Dienst benutzt.

[2] Aufteilung der Netzlast auf mehrere Server

[3] Funktionalität wird beim Ausfall einer Komponente durch eine andere übernommen. Dies soll möglichst so geschehen, dass der Anwender den Ausfall gar nicht bemerkt.

**Autorisierung**

1. Analog zur Re-Authentisierung muss auch die Re-Autorisierung nach einer definierten Dauer (oder vorher) durchgeführt werden.

2. unaufgefordertes Beenden: Über verschiedenen „Session Termination Messages“ kann der Client seinen Server infomieren, dass er die Sitzung beendet. Umgekehrt kann damit der Server auch den Client auffordern die Sitzung zu beenden.

**Abrechnen**

Das Ziel der Bestrebungen im Abrechnungsbereich ist es sicherlich den Verlust von Abrechnungsdaten unter den zahlreichen verschiedenen Fehlersituationen die auftreten können (Überlast, Failover, Ausfall einer Komponente, Übertragungsfehlern, etc.) zu minimieren oder bestenfalls ganz zu verhindern. Allerdings darf nicht nur auf die eigenen Bedürfnisse geachtet werden: Beispielsweise wäre es aus Sicht des „Abrechnens“ praktisch, jeden Client jede Sekunde auf seinen Zustand zu überprüfen. Dies würde allerdings zu keiner skalierbaren und ökonomisch sinnvollen Lösung führen. Deshalb wird die Übermittlung der Abrechnungsdaten entweder beim Eintreffen eines abrechnungsrelevanten Ereignis (Echtzeit-Modus) oder erst nach einer vordefinierten Dauer dafür mit allen bis dann angefallenen Ereignissen (Batch-Modus) durchgeführt. Es geht also hier um die Frage, was *Diameter* alles implementieren soll, um diesem Ziel gerecht werden zu können.

Hier eine Auflistung der wichtigsten Massnahmen:

1. Garantierte Aushändigung: Der *Diameter*-Server muss auf Applikationsebene explizit sämtliche Meldungen die mit einer verrechnenden Aktion zu tun haben (sprich, die fürs Accounting relevant sind) bestätigen.

2. Zeitstempel in jeder Abrechnungs-Nachricht

3. Dynamisches Abrechnen möglich: Durch das Anlegen von „Zwischen-Abrechnungs-Daten“, welche periodisch oder auch bei Re-Authentifizierung oder Re-Autorisierung versandt werden

Ein weiterer Punkt sollte hier noch Erwähnung finden, der zwar nicht dem oben genannten Ziel dient, sondern vielmehr die Attraktivität der Nutzung von *Diameter* verbessern kann: Die mögliche Erhöhung der Flexibilität im Zusammenhang mit den Abrechnungsdaten. Dies kann durch die Verwendung des einheitlichen Nachrichtenformat *ADIF* (Accounting Data Interchange Format) [2] erreicht werden; einem Standard der Syntax und Semantik von „Accountingdaten“ definiert und auch von *Diameter* verwendet werden kann.

### 10.2.3 Anforderungen im Zusammenhang mit Mobilität

Nachdem nun die Komponenten eines AAA-Prozesses, die Anforderungen an *Diameter* als AAA-Protokoll, sowie der Begriff „Mobilität“ in unserem Kontext betrachtet wurden,

werden wir in diesem Abschnitt den grössten Herausforderungen widmen, die sich aus dieser Sicht ergeben und deren Lösungsansätze betrachten.

1. Lokalisierung: Wie kann sichergestellt werden, dass ein mobiler Client immer gefunden werden kann? Die Netzwerkadresse, die bisher zur Identifizierung eines Clients reichte kann bei Wechsel von Netzwerken nicht beibehalten werden. Es müssen folglich Massnahmen getroffen werden, die dennoch eine Lokalisierung des Clients ermöglichen.

   *Diameter* verwendet dazu *Mobile IP*, welches unter Kapitel 10.4.1 näher erläutert wird.

2. Schutz sensilber Daten: So hoch der Nutzen sein mag, fremde Netzwerke für die Datenübertragung nutzen zu können, so birgt ein Netzwerk unter fremder Kontrolle/Verwaltung auch einige Gefahren. Sensible Daten sollen trotz der Übertragung durch das fremde Netzwerk nicht eingesehen werden können. Authentifizierungs-, Autorisierungs- und Abrechnungsdaten gehören ganz bestimmt auch in diese Kategorie. *Diameter* muss einen Weg finden, solche Informationen sicher übermitteln zu können.

   Dies geschieht über eine asymetrische Verschlüsselung.

3. Ausfallsicherheit: Da die verwendete Infrastruktur nicht unter eigener Kontrolle stehen muss, können Anforderungen an den Servicegrad nicht garantiert werden. Ein Ansatz wäre es, die Qualitätsanforderungen über das Protokoll auszuhandeln und dann ein Netzwerk zu wählen, welches den Anforderungen genügt. Ein anderer Ansatz liegt darin von einem nicht zuverlässigen Netzwerk auszugehen und im Protokoll selbst entsprechende Massnahmen zu implementieren, die für eine sichere Auslieferung/Verarbeitung nötig sind.

   *Diameter* verwendet TCP zur verlässlichen Übermittlung. Dass bei Systemausfällen oder anderen Fehlern keine Inkonsistenzen entstehen oder Abrechnungsdaten verloren gehen können, muss durch *Diameter* selber sichergestellt werden. Dies geschieht durch folgende Punkte:

   (a) Einführung von Bestätigungsmeldungen für abrechnungsspezifische Informationen: Sämtliche Meldungen müssen bestätigt werden (ACK-Meldung), bevor der AAA-Prozess fortgesetzt wird. Dadurch kann garantiert werden dass keine Prozesse in einen Zustand gelangen, den sie nicht haben dürften. Zum Beispiel soll keine Verrechnungsperiode begonnen werden, wenn die Gegenstation nicht deren Beginn bestätigt hat, weil beispielsweise die Verbindung abgebrochen ist.

   (b) Timeouts: Gewährleisten, dass Verbindungen oder Beziehungen nach einer definierten Zeitdauer ihre Gültigkeit verlieren und erneuert werden müssen. Fehlerhafte Situationen, wie beispielsweise der Absturz des „konsumierenden“ Endgerätes werden spätestens nach Ablauf solch eines Timeouts durchgeführt und damit die verrechnende Zeitdauer aller spätestens hier gestoppt. Eine zu kleine Timeoutzeit führt zu vielen Meldungen (und damit zur erneuten Authentifikation oder Autorisation, etc.) welche meist nicht nötig sind und somit zu einer „Verschwendung“ der Bandbreite, während eine zu grosse Timeoutzeit dazu führen kann, dass Fehler zu spät korrigiert werden was sich vorallem bei Fehlern

in den Abrechnungsdaten als ein eher ungünstiges Verhalten bezeichnen lässt. Üblich sind Timeoutzeiten von um die 60 Sekunden – je nach Situation.

4. Unterstützung durch bestehende Infrastruktur: Damit *Diameter* auch verwendet werden kann, müssen die bestehenden Netzgeräte dem Einsatz von *Diameter* auch ermöglichen. Sind viele Netzwerke vorhanden, auf denen dieses Protokoll keine Unterstützung findet, leidet auch seine Bedeutung (und damit auch die Chancen für eine Einsatz) entsprechend.

   *Diameter* ist nicht vollständig abwärtskompatibel zu seinem Vorgänger *RADIUS*. Allerdings wird ein sehr hoher Aufwand betrieben, um dennoch Interoperabilität zwischen beiden Protokollen (z.B. durch Gateways) sicherstellen zu können. Ein weiteres Problem stellen die Firewalls dar. Für eine zuverlässige Übertragung soll neben TCP das Protokoll SCTP eingesetzt werden, welches heute nicht von Firewalls erkannt wird. Die Pluspunkte von SCTP liegen in der Zuverlässigkeit und raschen Fehlererkennung. Beide kommen der Forderungen nach verlässlichen AAA-Meldungen nach. In Bezug auf das Protokoll SCTP gehen die Entwickler davon aus, dass es in der nächsten Zeit standardmässig von Firewalls unterstützt werden wird.

Eine weitere Frage, die weniger technischer Natur ist, sich aber meiner Meinung nach aus dieser Situation ergibt ist die der Transparenz der Kosten: Wann soll ein Netzwechsel zu welchem Provider stattfinden? Wer definiert nach welchen Kriterien (Preis, Qualität, Bandbreite) dies stattfinden soll? Soll dies auch automatisch geschehen oder explizit durch den Endanwender durchgeführt werden müssen (und damit die Transparenz der Netzwechsel verloren gehen)?

## 10.3 Diameter

### 10.3.1 Von *RADIUS* zu *Diameter*

*RADIUS* (Remote Authentication Dial In User Service) wurde 1992 von Steve Willens erstmals spezifiziert, um generell die Nutzung von AAA-Funktionalitäten zu ermöglichen. Um die Entwicklung eines offenen Standards zu unterstützen, gründete die IETF (Internet Engineering Task Force) 1995 die RADIUS Working Group, welche die grundlegenden Funktionalitäten und Formate von RADIUS [3] im RFC 2138 standardisierte.

*RADIUS* ist noch heute das AAA Protokoll mit der weltweiten grössten Verbreitung. Die grosse Akzeptanz und Verbreitung verdankt *RADIUS* seiner Herstellerunabhängigkeit. Im Gegensatz zu TACACS+ (Terminal Access Controller Access Control, AAA Protokoll von Cisco konzipiert) [4] und Kerberos [5] (AAA Protokoll von Merit konzipiert) wird *RADIUS* nicht von einem einzigen Hersteller entwickelt.

Ursprünglich wurde es für kleine Netzwerke mit wenigen Endnutzern entwickelt. Die Grundidee war ein Modem-Pool mit Wahlverbindungen nach aussen, die es zu überwachen galt.

Das konnte am effizientesten mit einer einzigen Datenbank mit Nutzerdaten zu Authentifizierung gewährleistet werden. Dadurch war keine aufwendige und teure Infrastruktur in den Modems selbst nötig.

### Eigenschaften *RADIUS*

1. Client-Server-Architektur

   Der NAS ist der Client bei *RADIUS*. Er ist für das Weiterleiten der Nutzerinformationen zum *RADIUS-Server* zuständig und muss auf die zurück gelieferten Antworten reagieren. Der *RADIUS-Server* erhält den Access-Request des Client und ist für dessen Authentifizierung und bei Erfolg für die notwendige Konfiguration verantwortlich. Der *RADIUS-Server* kann auch als Proxy-Client zu anderen *RADIUS*- oder Authentifizierungs-Servern dienen (siehe Abbildung 10.1).

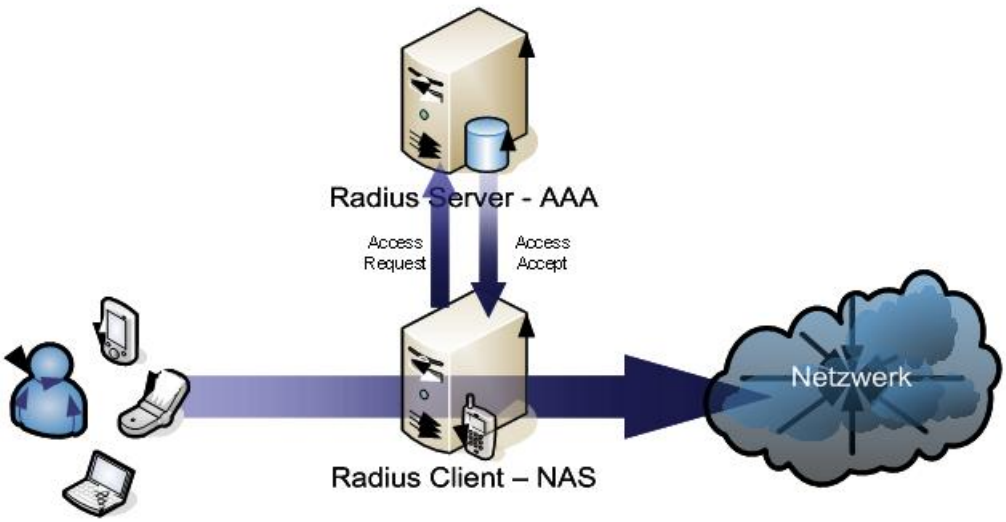


Abbildung 10.1: RADIUS Architektur

2. Netzwerk Sicherheit

   Die zwischen RADIUS Client und Server ausgetauschten Nachrichten werden mittels eines gemeinsamen Schlüssels, „Shared Secrets“, authentifiziert, der niemals über das Netzwerk ausgetauscht wird. Zusätzlich werden Benutzer-Passwörter für die Übertragung verschlüsselt.

3. Flexibler Authentifizierungsmechanismus

   RADIUS Server können eine Vielzahl von Methoden zur Authentifizierung eines Users unterstützen. Zu nennen sind unter anderem PAP (Password Authentication Protocol), CHAP (Challenge Handshake Authentication Protocol) oder EAP (Extensible Authentication Protocol).

4. Erweiterbares Protokoll – Attribute Value Pairs

RADIUS Nachrichten bestehen aus AAA Informationen, welche in den so genannten „Attribute Value Pairs“ mit der Tupellänge 3 gespeichert werden. Typischerweise enthalten diese Attribute den User-Name, das User-Passwort, das gekapselte Protokoll, die gekapselten Daten etc. Neue Attribute können, ohne die vorhandene Implementierung des Protokolls abzuändern, hinzugefügt werden.

### Protokollablauf RADIUS

Der Authentifizierungsprozess in einem Netzwerk findet unter Einsatz von RADIUS im Wesentlichen wie folgt statt:

Ein Enduser wählt sich bei einem NAS ein, bei dem ein RADIUS-Client installiert ist. Der NAS ermittelt, z.B. PPP-Frames, den Usernamen und das Passwort des Endusers. Danach sendet der NAS mittels UDP/IP die verschlüsselte „Access-Request“ Nachricht zum RADIUS Server. Diese Nachricht kann zusätzliche Attribute beinhalten, z.B. NAS Port und IP Adresse.

Der RADIUS Server sucht nun in seiner Datenbank nach einem Eintrag mit dem entsprechenden Username. Existiert kein entsprechender Eintrag, sendet der RADIUS-Server eine „Access-Reject“ Nachricht an den NAS, die optional eine Text Nachricht mit der Fehlerursache enthält. Der NAS benachrichtigt jetzt den Enduser, dass ein Fehler während der Authentifizierung aufgetreten ist.

Wenn ein Eintrag für den User gefunden wird und das gelieferte Passwort richtig ist, sendet der RADIUS Server eine „Access-Accept“ Nachricht an den NAS. Darüber hinaus sendet er zusätzliche Konfigurationsdaten an den NAS. Diese Konfigurationsdaten werden benötigt um die Netzwerkverbindung herzustellen und beinhalten z.B. eine IP Adresse für den End User oder einen Filter, der den Benutzer auf die Verwendung bestimmter Protokoll-Typen beschränkt (z.B. Telnet oder HTTP).

### RADIUS Protokoll

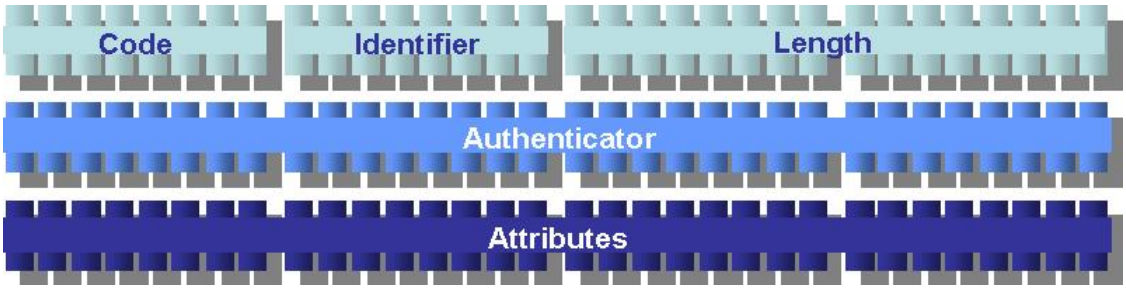


Abbildung 10.2: RADIUS Header

RADIUS sendet Informationen mittels RADIUS-Paketen, welche aus einem Header (siehe Abbildung 10.2) und verschiedenen RADIUS-Objekten bestehen. Diese Objekte, die Attribute Value Pairs, bestehen aus einem Header und Daten.

Der Header eines RADIUS Paketes besteht aus 4 verschiedenen Feldern:

1. Ein 8 Bit Codefeld, welches den Typ des Pakets enthält.
2. Ein 8 Bit Kennungsfeld (Identifier), welches die Verknüpfung von Anfragen und Antworten ermöglicht.
3. Ein 16 Bit Längenfeld (Length), welches die Länge des Pakets enthält, inklusive der Länge aller AVP's.
4. Ein 128 Bit Authentifikationsfeld (Authenticator), welches zur Authentifizierung der Antwort des RADIUS Servers genutzt wird. Darüber hinaus wird es zur Übertragung des verschlüsselten Client Passworts bei der Anfrage genutzt.

Die RADIUS-AVP's beinhalten die spezifischen Authentifizierungs-, Autorisierungs- und Accountinginformationen und Konfigurationsdetails bezüglich der einzelnen Requests und Responses.

### Schwächen von RADIUS

Nachdem RADIUS-Implementierungen seit über zehn Jahren im Einsatz sind, zeigen sich Schwächen des Protokolls, die eine Nutzung unter gestiegenen Anforderungen in punkto Transportsicherheit, Skalierbarkeit und Flexibilität in Frage stellen.

Der Datentransport per UDP stellt sich als Schwachstelle heraus. Sehr wohl beschreibt RADIUS eine Strategie, wie auf ausbleibende Antworten eines Servers zu reagieren ist, nämlich mit Wiederholungsversuchen nach einem Timeout. Doch die Anzahl der Wiederholungen als auch die Zeitintervalle sind nicht im Protokoll festgelegt, so dass es zwischen verschiedenen Implementierungen zu Inkompatibilitäten kommen kann.

Der Server sendet dem Client keine Status- oder Fehlermeldungen. Somit ist weder in der Transportschicht noch auf höherer Ebene eine Fehlererkennung implementiert. Der Client kann lediglich anhand der Timeouts erkennen, ob der Server erreichbar ist. Muss der Server zur Bearbeitung der Anforderungen jedoch auf langsame Legacy-Systeme zugreifen, so interpretiert der Client eine lange Antwortzeit eventuell als einen Fehler während des Datentransports oder des Servers und reagiert unangemessen mit einer Wiederholung der Anfrage.

Tritt nun tatsächlich ein Übertragungsfehler zwischen einem Proxy und einem Server auf, so reagiert nicht der Proxy mit einem neuen Versuch, sondern der NAS beziehungsweise der Client. Dabei wird ein unnötig hoher Datenverkehr in Kauf genommen.

Da RADIUS einem Client-Server-Modell entspricht, ist keine Möglichkeit vorgesehen, dass der Server eine Kommunikation mit dem Client beginnt. Besonders im Bereich Accounting ist dieses Verfahren aber üblich, um Informationen in Echtzeit anzufordern oder den Client aufzufordern, die Session eines Nutzers zu beenden.

Die Daten in den Nachrichten sind byteweise ausgerichtet. Da aktuelle Rechnerarchitekturen nur auf Daten, die an 32 Bit-Grenzen ausgerichtet sind, besonders schnell zugreifen, verschenkt man hier Leistung.

RADIUS ist nicht modular erweiterbar. Zwar lassen sich weitere Attribute definieren, doch insgesamt sind nur 256 verschiedene erlaubt. Implementieren nun verschiedene Hersteller neue Funktionen und Attribute, so kann es zu Überschneidungen in dem begrenzten Namensraum und damit Inkompatibilitäten kommen. Außerdem ist die Länge eines AVP's durch das lediglich ein Byte grosses Längenfeld auf 256 Bytes beschränkt. Sollen grössere Datenmengen versendet werden, so müssen sie auf mehrere AVP's verteilt werden.

#### Diameter – eine Erweiterung von RADIUS

Mit der Zunahme der Komplexität der geforderten Dienste, der Anzahl Router und der zunehmenden Dichte von NAS erweist sich RADIUS als ungeeignet zur Nutzung in grösseren Netzwerken. Insbesondere hat sich gezeigt, dass Roaming-Vereinbarungen mit allen ISP's untereinander nicht skalieren. Dies machte die Entwicklung eines neuen Protokolls notwendig.

Um den Problemen des RADIUS Protokolls zu begegnen, wurde schon 1996 von Pat Calhoun mit der Entwicklung von Diameter begonnen. Zusammen mit Firmen wie Bell, Cisco oder Sun Mircrosystems, war das Ziel, einen Nachfolger für RADIUS zu finden. Dabei wurde besonders auf Kompatibilität mit RADIUS Wert gelegt, um eine schrittweise Migration zu Diameter zu erleichtern.

Der Name Diameter wurde als Pendant zu RADIUS gewählt, da der Durchmesser eines Objekts (auf Englisch „Diameter“) symbolisch eine Erweiterung/Verlängerung des RADIUS ist. Somit steht Diameter nicht wie RADIUS für eine Abkürzung.

### 10.3.2 Diameter Architektur

Diameter ist in einen allgemeinen Teil, das Base Protocol [14], und in anwendungsspezifische Erweiterungen, den so genannten Applications, unterteilt. Das Base Protocol beschreibt den Transport der Daten, Fehlermeldungen und grundlegende AAA-Funktionalitäten sowie dem Sessionmanagement. Die Applications stellen dann die Unterstützung für bestimmte Technologien, wie Mobile IP oder Dial-In Zugänge, zur Verfügung. Durch das Konzept der Applications ist Diameter modular erweiterbar und besitzt eine gewisse Flexibilität, um auf Neuerungen in der Technologie reagieren zu können. Gegenwärtig sind jedoch nur Applications für Mobile IPv4 und Network Access Server spezifiziert.

Abbildung 10.3 zeigt eine schematische Darstellung der Diameter Architektur. Das Basis Protokoll ist eng verbunden mit der CMS (Cryptographic Message Syntax) Applikation, um Sicherheit für alle Applikationen zu gewährleisten. Alle Applikationen müssen hierbei, trotz unterschiedlicher Funktionalitäten, dass Diameter Basis Protokoll unterstützen.

Statt einer Client-Server-Architektur, wie wir sie bei RADIUS antreffen, wurde bei Diameter eine Peer-to-Peer-Architektur konzipiert. Der Server kann den Client unaufgefordert Anfragen schicken, ohne erst ein request vom Client zu erhalten, wie es typischerweise in

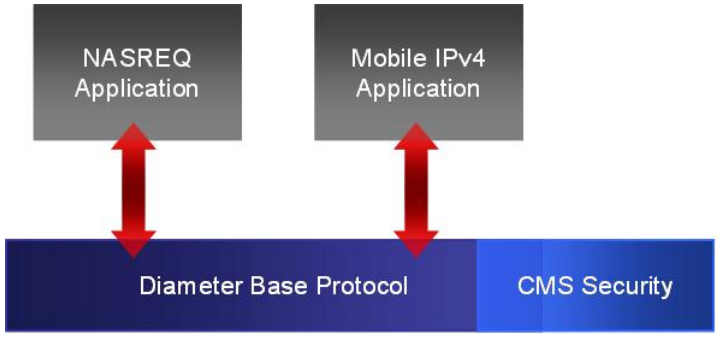


Abbildung 10.3: Diameter Architektur

ein Client-Server-Umgebung geschieht. Ein solches Prinzip nennt man „Server-directed-Model“. Vor allem im Hinblick auf die Accounting-Funktionalitäten ist eine solche Architektur von grossem Nutzen. Der Server kann vom Client Accounting-Daten anfordern, das Terminieren von Sitzungen fordern oder die Sessions allgemein überwachen.

### 10.3.3 Diameter Komponenten

Zusätzlich zu den Clients und Server findet man bei Diameter zusätzliche Agenten, die in verschiedenster Hinsicht sehr nützlich sind. Hier eine Übersicht zu den verwendeten Abkürzungen:

- NAS: Network Access Server
- DRL: Diameter Relay (Agent)
- HMS: (Diameter) Homer Server
- DRD: Diameter Redirect Agent
- TLA: Diameter Translation Agent

#### RelayAgents

Relay Agents leiten Nachrichten anhand von Routing AVP’s in den Nachrichten und Routingtabellen weiter. Dabei können Relay Agents der Nachricht lediglich Routing-Informationen hinzufügen und arbeiten sonst transparent. Andere Informationen werden nicht modifiziert.

#### Proxy Agents

Ähnlich wie die Relay Agent leiten die Proxy Agent ebenso Nachrichten weiter, können jedoch ungültige Anfragen mit Fehlermeldungen und Access-Reject-Meldungen abfangen. Proxies müssen also die Nachrichten auswerten und damit verwendetete Applications

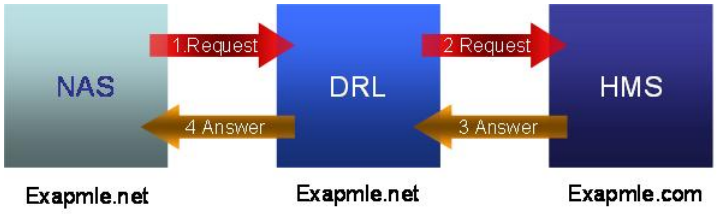


Abbildung 10.4: Relay Agent Diameter

auch unterstützen. Vor allem für ISP's wären Proxy Agents sehr nützlich. Sie können die Anzahl und die Typen der verwendeten Ports überwachen, können Allokations- und Zugangsentscheidungen treffen.

### Redirect Agents

Redirect Agents senden dem Client auf eine Anfrage hin die Adresse des Zielservers, so dass eine direkte Kommunikation zwischen Client und gesuchtem Server möglich ist.

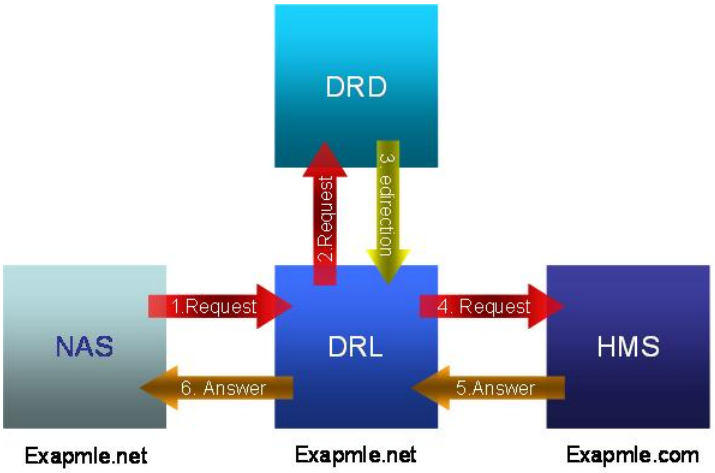


Abbildung 10.5: Redirect Agent Diameter

### Translation Agents

Translation Agents stellen Gateways dar, die AAA-Protokollnachrichten übersetzen, typischerweise zwischen RADIUS- und Diameter-Implementierungen. So wird die Migration auf Diameter vereinfacht, die Server können schnell getauscht werden und die grosse Zahl an Clients muss nicht sofort umgerüstet werden.

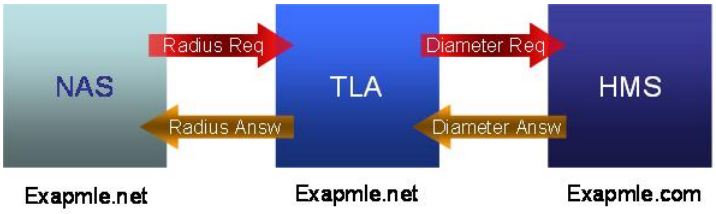


Abbildung 10.6: Translation Agent Diameter

### 10.3.4 Diameter Header

Diameter-Nachrichten sind bit-orientiert aufgebaut, sie bestehen aus einem statischen Kopf, der allgemeine Informationen über die Nachricht beinhaltet und „Attribute Value Pairs“, welche die eigentlichen Nutzdaten enthalten.

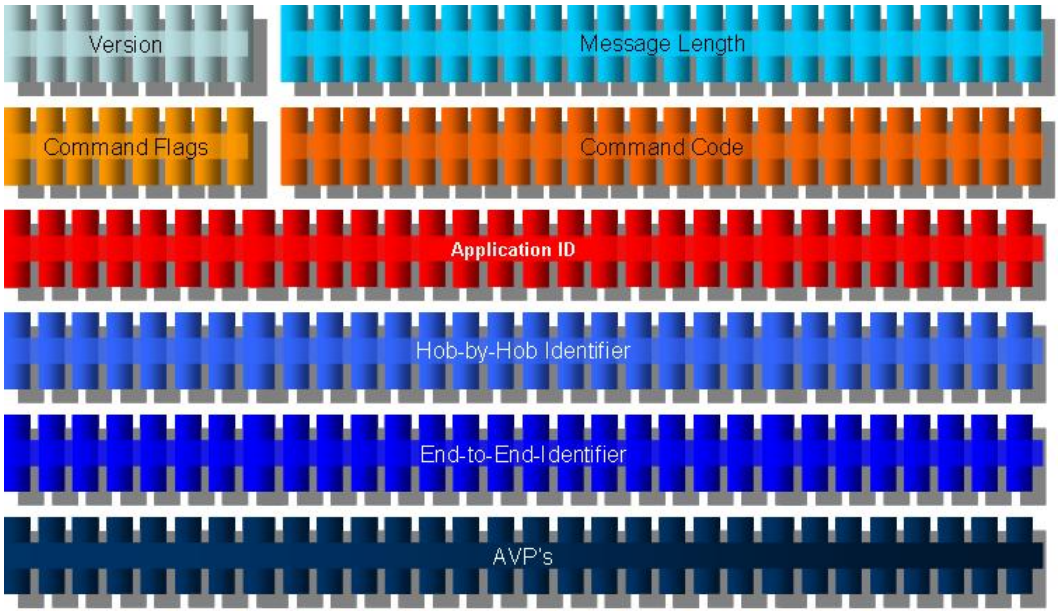


Abbildung 10.7: Diameter Header

1. Version: Dieses Feld ist auf 8 Bit begrenzt und zeigt die Diameter-Version an. Aktuell ist dies die Version 1.

2. Message Length: Hier wird die Länge der Nachricht inklusive des Headers in 24 Bit festgehalten.

3. Command Flags: Dieses Feld besteht aus 8 Bit, wobei es folgendermaßen belegt ist:

   (a) Das R-Bit wird gesetzt, wenn die Nachricht eine Anfrage (Request) ist. Falls das R-Bit nicht gesetzt ist, handelt es sich um eine Anwort.

   (b) Das P-Bit ist gesetzt, wenn die Nachricht über Proxys weitergeleitet werden darf.

   (c) Das E-Bit wird bei Fehlernachrichten gesetzt.

Tabelle 10.1: Nachrichtenmethoden Diameter

| *Name* | *Command-Code* | *Beschreibung* |
|---|---|---|
| Capabilities-Exchange-Request / Answer (CER, CEA) | 257 | Austausch der Fähigkeiten zwischen zwei Knoten |
| Device-Watchdog-Request / Answer (DWR, DWA) | 280 | Überwachung der Transportverbindung |
| Disconnect-Peer-Request / Answer (DPR, DPA) | 282 | Abbau der Transportverbindung |
| Abort-Session-Request / Answer (ASR, ASA) | 274 | Abbruch einer Session |
| Accounting-Request / Answer (ACR, ACA) | 271 | Übermittlung von Accounting-Informationen |
| Re-Auth-Request / Answer (RAR, RAA) | 258 | Erneute Authentisierung |
| Session-Termination-Request / Answer (STR, STA) | 275 | Beendigung eine Session |

(d) Das T-Bit wird nach einer Verbindungsunterbrechung gesetzt und bedeutet, dass diese Nachricht möglicherweise doppelt gesendet worden ist. Dies ist hilfreich, um doppelte Nachrichten zu beseitigen.

(e) Die r-Bits sind zukünftigen Benutzungen vorbehalten.

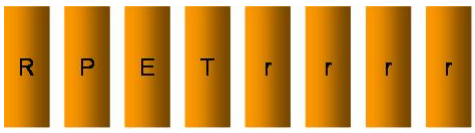


Abbildung 10.8: Command Flags Diameter Header

4. Command-Code: Dieses 24-Bit-Feld beschreibt die Methode der Nachricht. Die Tabelle 10.1 enthält eine Übersicht der existierenden Nachrichtenarten im Basisprotokoll.

5. Application-Id: Die Application ID identifiziert den Typ der Applikation. Dabei kann es sich um eine AAA-Applikation, oder einer händlerspezifischen Applikation handeln.

6. Hop-By-Hop Identifier: Dieses Feld hat eine Länge von 32 Bit und wird von den Diameter-Proxys zur Zuordnung der Diameter-Nachrichten verwendet.

7. End-To-End Identifier: Dieses Feld hat eine Länge von 32 Bit und dient dem Sender und Empfänger der Nachricht zur korrekten Zuordnung.

### 10.3.5 AVP Header

In den Attribute Value Pairs werden die eigentlichen Informationen übertragen. Dabei sind je nach Nachrichtentyp unterschiedliche AVP's vorhanden. Alle AVP's verfügen über

den identischen Headeraufbau.

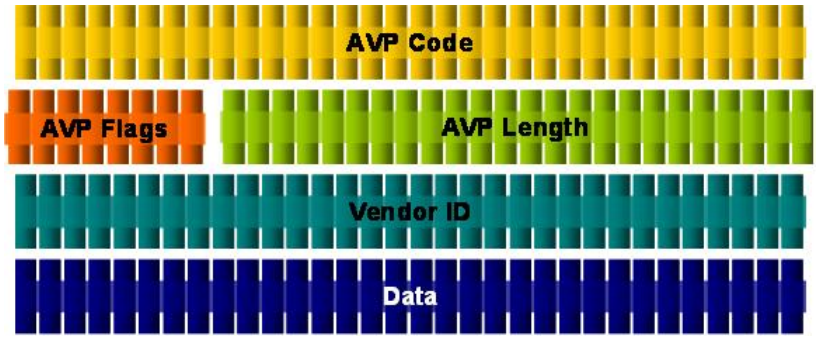


Abbildung 10.9: AVP Header

1. AVP Code: Der 32 Bit lange AVP Code zusammen mit der optionalen Vendor-Id identifiziert das AVP eindeutig. Der Bereich 0 bis 255 ist für die Abwärtskompatibliltät mit RADIUS reserviert, der Bereich darüber spezifiziert die Diameter-AVPs.
2. AVP Flags: Dieses Feld besteht aus 8 Bit und enthält Informationen über das AVP:

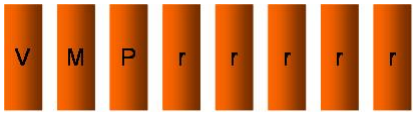


Abbildung 10.10: AVP Flags AVP Header

   (a) Das V-Bit wird gesetzt, wenn es sich um ein herstellerspezifisches AVP handelt. Nur wenn dieses Bit gesetzt ist, existiert das Vendor-Id Feld im AVP-Header.
   (b) Ein gesetztes M-Bit zeigt an, dass dieses AVP für die aktuelle Nachricht notwendig ist. Dieses Bit ist für das Parsen der Nachricht wichtig, denn falls es nicht gesetzt und das AVP somit optional ist, muss der Empfänger der Nachricht dieses AVP nicht unbedingt kennen, um die Nachricht verarbeiten zu können.
   (c) Das P-Bit beschreibt die Notwendigkeit einer Verschlüsselung für dieses AVP.
   (d) Die r-Bits sind für zukünftige Anwendungen vorbehalten.
3. AVP Length: Das 24 Bit grosse Feld beschreibt die Länge des AVP's inklusive des AVP-Headers.
4. Data: AVP's können unterschiedliche Typen von Daten transportieren. Eine Übersicht ist der Tabelle 10.2 zu entnehmen.

### 10.3.6 Wichtige AVP's

Es folgt nun eine Beschreibung der wichtigsten AVP's, die unter anderem in Diameter Base Protocol standardisiert sind:

Tabelle 10.2: AVP Basistypen

| *Basistyp* | *Zulässige AVP Länge in Bytes* | *Beschreibung* |
|---|---|---|
| OctetString | minimal 8 (12 bei gesetztem V-Bit) - muss auf ein Vielfaches von 32 Bit enden | Dient zum Transport von Texten. Falls der String nicht auf 32 Bit endet, muss bis zur nächsten 32 Bit Grenze aufgefüllt werden |
| Unsigned32 | 12 (bei gesetztem V-Bit 16) | Eine 32 Bit Integer-Zahl ohne Vorzeichen |
| Unsigned64 | 16 (bei gesetztem V-Bit 20) | Eine 64 Bit Integer-Zahl ohne Vorzeichen |
| Float32 | 12 (bei gesetztem V-Bit 16) | Eine 32 Bit Zahl mit Nachkommastellen |
| Float64 | 16 (bei gesetztem V-Bit 20) | Eine 64 Bit Zahl mit Nachkommastellen |
| Grouped | 8 (12 bei gesetztem V-Bit) + die Länge aller in der Gruppe vorhandenen AVP's | Ein grouped AVP fasst unterschiedliche AVP's zu einem AVP zusammen |

**Session-Id:**

Dieses AVP ist vom Typ UTF8String, hat den Code 263 und dient zur Identifizierung einer spezifischer Session. Die Session-Id wird von dem Diameter-Node generiert, der die Session erstellt, wobei dies normalerweise der Client ist.

**Origin-Host:**

Das Origin-Host AVP ist vom Typ DiameterIdentity, hat den Code 264 und ist in allen Diameter-Nachrichten vorhanden. Durch dieses AVP kann der Ursprung der Nachrichtnachvollzogen werden.

**Accounting-Record-Type:**

Dieses AVP ist vom Typ Enumerated und hat den Code 480. Es ist in allen Accounting-Anfragen vorhanden und beinhaltet den Typ der Accounting-Anfrage:

1. `EVENT_RECORD(1)`: Dieser Wert wird bei einmaligen Ereignissen genutzt.

2. `START_RECORD(2)`: Dieser Wert initiiert eine Accounting Session, die sich auf einen Dienst mit messbarem Wert bezieht.

3. `INTERIM_RECORD(3)`: Dieser Wert wird gesetzt, wenn während einer auf einen Dienst mit einem messbaren Wert bezogenen Accounting Session eine Accounting Nachricht gesendet wird. Dies geschieht beispielsweise nach einer Re-Authentisierung.

4. `STOP_RECORD(4)`: Dieser Wert beendet eine Accounting Session, die sich auf einen Dienst mit einem messbaren Wert bezieht.

**Accounting-Record-Number:**

Dieses AVP ist vom Typ Unsigned32, hat den Code 485 und identifiziert einen Accounting-Record. In Kombination mit der Session-Id kann ein Accounting-Record eindeutig nachvollzogen werden.

**Authorization-Lifetime:**

Dieses AVP ist vom Typ Unsigned32, hat den Code 291 und enthält die maximale Anzahl der Sekunden, in denen ein Nutzer auf einen Dienst zugreifen darf, ohne sich reautorisieren zu müssen.

**Auth-Grace-Period:**

Dieses AVP ist vom Typ Unsigned32, hat den Code 276 und beinhaltet die Anzahl der Sekunden die ein Server nach Ablauf der Authorization-Lifetime wartet bis er die Ressourcen der Session entfernt.

**Auth-Session-State:**

Dieses AVP ist vom Typ Enumerated und hat den Code 277. Es spezifiziert, ob der Zustand einer Session gespeichert wird:

1. `STATE_MAINTAINED`: Dieser Wert bedeutet, dass der Zustand gespeichert wird.

2. `NO_STATE_MAINTAINED`: Dieser Wert bedeutet, dass der Zustand nicht gespeichert wird.

**Accounting-Realtime-Required:**

Dieses AVP ist vom Typ Enumerated und hat den Code 483. Es enthält Informationen für einen Diameter-Client bezüglich des Umgangs mit anfallenden Accounting-Daten, falls die Verbindung zwischen dem Diameter-Client und Diameter-Server beispielsweise durch Netzwerkprobleme unterbrochen sind:

1. `DELIVER_AND_GRANT`: Dieser Wert bedeutet, dass der Dienst nur solange bereitgestellt werden darf, wie es eine Verbindung zwischen dem Client und dem Server gibt.

2. `GRANT_AND_STORE`: Dieser Wert bedeutet, dass der Dienst nur solange bereitgestellt werden darf, wie es eine Verbindung zwischen dem Client und dem Server gibt, oder die Accounting-Daten beim Client zwischengespeichert werden können.

3. `GRANT_AND_LOSE`: Dieser Wert bedeutet, dass der Dienst auch noch bereitgestellt werden soll, wenn die Accounting-Daten nicht gesendet und nicht zwischengespeichert werden können.

### 10.3.7 Sicherheit

#### Transportsicherheit

Die Übertragung der Daten geschieht via das zuverlässige, verbindungsorientiert TCP, und nicht wie bei RADIUS über UDP. Somit ist auf Transportebene Flusskontrolle und Fehlervermeidung gewährleistet.

SCTP, ein jüngeres Protokoll, welches die Funktionalität von TCP erweitert, und vor allem auf Bedürfnisse der heutigen komplexen Kommunikation zugeschnitten wurde. Im Gegensatz zu TCP sind zwischen zwei Kommunikationspartnern nicht nur ein Datenstrom möglich, sondern gleichzeitig mehrere Datenströme. In jedem Datenstrom werden die Daten jeweils separat in der richtigen Reihenfolge übertragen. Somit hat ein Fehler in einem Datenstrom, zum Beispiel durch Abbruch einer Verbindung oder einer Verzögerung der Übermittlung, keinen Einfluss auf die restlichen Datenströme.

Fehlererkennung ist aber auch auf Anwendungsebene möglich. Jede Anfrage eines *Nodes* muss vom Server beantwortet werden. Die *Response* enthält einen „Result-Code-AVP“, in dem ein Fehler- oder ein Erfolgstatus kommuniziert wird.

Zusätzlich zu den oben genannten Massnahmen wird die Sicherheit durch folgende Punkte erhöht:

1. Hob-by-Hop Security

   Jede Diameter Application muss IPsec als Sicherheitsmechanismus unterstützten. Daneben kann optional noch Transport Layer Security (TLS, auch bekannt unter dem Namen „Secure Sockets Layer“; SSL) genutzt werden.

2. Diameter Path Authorization

   Zusätzlich zur Transportsicherheit definiert Diameter eine Autorisation einer Session als ganzes. Bevor überhaupt eine Verbindung aufgebaut wird, überprüft ein Diameter peer ob die anderen peers berechtigt sind, in ihren Rollen zu agieren (Relay Agent, Translation Agent usw.) So kann zum Beispiel ein peer authentisch sein, aber er ist nicht berechtigt als Server zu agieren.

Der Home-Diameter-Server überprüft die Route-Recod-AVP's, welche von den Proxys erstellt wurden, bevor er die Session autorisiert.

#### Ausfallsicherheit

Durch das Versenden von „Device-Watchdog-Request"- und „Device-Watchdog-Answer"-Nachrichten kann proaktiv auf Ausfälle reagiert werden, wenn zum Beispiel ein peer idle ist, oder wenn eine Answer auf ein Request nicht innerhalb der Zeitlimite ankommt.

Wenn nun doch ein Ausfall eines peers entdeckt wird, sendet der Diameter-Node alle Daten zu einem Alternativ-Peer. Der Diameter Node versucht gleichzeitig, in periodischen Abständen Verbindung zum ausgefallenen Peer aufzubauen. Kann die Verbindung wieder hergestellt werden, so leitet der Node alle Messages erneut weiter. Dieses Szenario nennt man „Failover/Failback"-Prozess

#### Ende-zu-Ende Sicherheit

Die sichere Übertragung auf Transportschicht bzw. Internetschicht wird mittels TLS bzw. IPsec realisiert. Wenn aber zwischen zwei Peers ein oder mehrere Agents liegen, ist eine Absicherung der Nachrichten über IPsec oder TLS nicht geeignet, da die Nachrichten bei jedem Agent auf dem Anwendungsschicht verarbeitet werden.

Mittels der CMS Security Application kann eine Ende-zu-Ende-Authentifikation, Integrität und Vertraulichkeit der Daten auf AVP Level gewährleistet werden. CMS verwendet dafür folgende Methoden: Digitale Signaturen und Zertifikate, sowie ein Verschlüsselungsalgorithmus.

### 10.3.8 Vergleich RADIUS und Diameter

Nach der Darstellung der Grundkonzeption der beiden AAA-Protokolle sollen die Unterschiede zwischen den beiden in Bezug auf Erweiterbarkeit, Verlässlichkeit, Flusskontrolle und Skalierbarkeit zusammengefasst werden.

Neben der unterschiedlichen Architektur unterscheidet zuerst die Erweiterbarkeit die beiden Protokolle: Während RADIUS nur einen begrenzten Befehl- und Attributplatz von maximal 256 Attributen bietet und deswegen als nicht sehr erweiterbar anzusehen ist, wird dieses Problem bei Diameter mittels eines erweiterbaren Basis Protokolls und 32 Bit Attributen gelöst. Jeder denkbare neue Dienst kann in Diameter durch die Erweiterung des Basis Protokolls implementiert werden, ohne die Basis ändern zu müssen.

Zuverlässigkeit beschäftigt sich mit Fragen zur Zustellung von Nachrichten zwischen den Netzwerkelementen und Flusskontrolle. RADIUS setzt, wie oben erklärt, UDP zum Datentransport zwischen Client und Server ein. Zusätzlich ist eine Time-out- und Übertragungswiederholungsstrategie implementiert. Jedoch wurde kein Standardschema definiert, was zu verschiedenen Implementierungen und daraus resultierenden Problemen geführt hat.

Diameter arbeitet mit TCP und bietet sowohl Flusskontrolle wie auch Congestion Avoidance an. RADIUS implementiert keine Flusskontrolle über UDP.

Skalierbarkeit ist im Hinblick auf die zu erwartenden Nutzerzahlen der AAA-Dienste sehr wichtig. Zu beachten sind Szenarien, in denen viele Nutzer gleichzeitig AAA-Funktionalitäten beim gleichen Server nutzen wollen.

Bezüglich implementationsspezifischen Gesichtspunkten ergibt sich folgender Unterschied: RADIUS Nachrichten werden byteweise angeordnet, während sie bei Diameter 32-bit angeordnet sind. Dies ermöglicht eine höhere Zahl von Transaktion pro Sekunde pro Server bei Diameter.

Das RADIUS Protokoll bietet nur eine hop-by-hop-Sicherheit an und hat keinen Sicherheitsmechanismus implementiert, um die Daten auf AVP-Ebene zu verschlüsseln. Diese Sicherheitslücke könnte es Proxy-Servern ermöglichen, vertrauliche Daten zu sammeln oder zu modifizieren (z.B. Accounting-Informationen), sodass End-Nodes dies nicht bemerken würden. Diameter bietet zusätzlich zur hop-by-hop-Sicherheit noch eine End-to-End-Sicherheit. Durch digitale Signaturen ist die Integrität der AVP's gewährleistet und mittels einem Verschlüsselungsalgorithmus kann man die Vertraulichkeit der Daten garantieren.

## 10.4 AAA im Mobilen Bereich

Will das Endgerät nicht nur innerhalb eines Netzwerkes sondern auch zwischen verschiedenen Netzen mobil sein, ergeben sich dadurch einige Probleme, die es zu lösen gilt. In diesem Kapitel wollen wir Eigenschaften identifizieren, die *Diameter* ausweisen muss, sollte es sich in einer mobilen Anwendungsumgebung behaupten wollen. Ein wichtiges Konzept von *Diameter* ist seine Modularität. Eine dieser Erweiterungen, diejenige für *Mobile IP* [10] soll anschliessend als Beispiel dienen, wie der konkrete Einsatz in einem mobilen Umfeld aussehen soll.

Zunächst wollen wir aber beispielhaft erläutern, was wir in unserer Arbeit unter dem Wort „Mobilität“ verstehen. Sie lässt sich grob in drei Szenarien unterteilen:

1. Mobilität des Endanwenders: Der Endanwender hat von verschiedenen Standorten aus über die jeweiligen vorhandenen Endgeräte Zugriff auf seine vom Datenserver verwaltete Dateien.

2. Mobilität der Endgeräte: Ähnlich zum oben genannten Fall besitzt der Endanwender mit seinem eigenen Endgerät von jedem beliebigen Netzwerk her Zugriff auf den Datenserver. Beispiele für derzeit übliche mobile Endgeräte sind *Laptops*, *Mobiltelefone* und *PDA's* (Personal Digital Assistant; ein meist handliches Endgerät für die Verwaltung persönlicher Informationen wie Adressen, Termine, etc.)

3. Mobilität der Nutzung: Der Benutzer bewegt sich während der Nutzung eines Dienstes mit seinem Endgerät von einem Netzwerk in beliebig viele andere ohne den Zugriff auf seinen Datenserver zu verlieren.

Bei allen oben genannten Punkten handelt es sich nicht um eine strenge Definition, sondern vielmehr um eine beispielhafte Beschreibung der wesentlichen Eigenschaften. Natürlich kann der Zugriff auf den Datenserver als der Zugriff auf ein beliebiges, „berechtigtes“ und „erreichbares“ Netz verallgemeinert werden, wobei wir aber zu präzisieren haben, was nun genau ein „berechtigtes“ Netzwerk ist und was ein „Erreichbares“. Beispielsweise soll nur eine sehr stark begrenzte Zahl von Personen einen *berechtigten* Zugriff auf ein Netzwerk verfügen, über welches die Steuerung eines Atomkraftwerkes (AKW) stattfindet. Idealerweise ist solch ein sensibles Netz isoliert und nicht über ein anderes Netzwerk wie z.B. dem Internet *erreichbar*. Der Zugriff auf den Datenserver fasst somit beispielhaft zusammen, dass ein **gültiger Zugriff** auf ein anderes Netzwerk stattfindet. Findet kein solcher Zugriff statt, wird in der Regel dennoch von Mobilität gesprochen. In unserem Kontext interessieren uns aber nur die oben genannten Fälle, wo ein Zugriff auf ein anderes Netzwerk stattfinden soll. Gerade dieses Problem ist ja Gegenstand unserer Arbeit. Wenn wir also von Mobilität sprechen meinen wir eine der drei oben genannten Situationen.

Was bedeutet dies nun für ein AAA-Protokoll? Was muss alles gewährleistet werden, um solch eine „Mobilität“ durch das Protokoll unterstützen zu können? Betrachten wir nochmals die entsprechenden Situationen:

1. Mobilität des Endanwenders: Das Endgerät muss ein AAA-Prozess unterstützen und über eine Verbindung zum Provider verfügen, bei welchem der Endanwender registriert ist. Daneben muss das eigene AAA-Protokoll identisch oder mindestens kompatibel zu demjenigen des Providers sein.

2. Mobilität der Endgeräte: Dieser Fall ist von Natur aus sehr mit der *Mobilität des Endanwenders* verwandt, da die Nutzung mobiler Endgeräte auch einen mobilen Endanwender mit einschliesst. Allerdings können unter Umständen Teile des AAA-Prozesses anders verlaufen (oder sogar wegfallen), wenn davon ausgegangen wird, dass **nur** der Endanwender Zugriff auf seinem Endgerät hat[4].

3. Mobilität der Nutzung: Auch hier muss selbstverständlich die nötige Grundvoraussetzung im Vorhandensein der Infrastruktur sowie entsprechender Protokolle erfüllt sein. Die unterbruchslose Nutzung in verschiedenen Netzen stellt hier die schwierigste Anforderung dar. Um einen Netzwechsel erfolgreich vollziehen zu können, müssen verschiedene Massnahmen getroffen werden, um die Adresse des Endgerätes zu wechseln, die AAA-Informationen sicher auszutauschen sowie eine Neuregistrierung vorzunehmen ohne dass die Verbindung bzw. der Prozess unterbrochen wird. Falls möglich sollte auch hier die Transparenz bewahrt bleiben, so dass dem Endanwender der Netzwechsel verborgen bleibt.

Das dritte Szenario erscheint nicht nur relevant, es beinhaltet auch die beiden anderen. Unter Kapitel 10.4.1 wird die Umsetzung dieser Forderungen durch *Diameter* und *Mobile IP* beschrieben.

[4]Was in der Realität zwar oft beabsichtigt, aber nur äusserst selten sicher realisierbar ist

### 10.4.1 *Diameter* und Mobile IP

*Diameter* für sich selbst (standalone) kann von den AAA-Aufgaben eigentlich nur die „Abrechnung“ bewältigen. Allerdings wurde es bewusst so konzipiert, einfach erweiterbar zu sein. Heute existieren zwei Erweiterungen für *Diameter*, welche Authentifizierung und Autorisierung zusätzlich ermöglichen: *NASREQ* [9] und *Mobile IP* [10]. Letztes soll nun als Beispiel für den mobilen Einsatz von *Diameter* dienen.

#### Mobile IP

Um das Problem der Lokalisierung in den Griff zu bekommen hat man sich bei *Mobile IP* dazu entschieden einen sogenannten *Home Agent* zu verwenden, der den Standort der mobilen Hosts seines Netzwerks verwaltet und Daten an einen Host weiterleitet, falls sich dieser nicht im eigenen Netz befindet. Wechselt ein Host in ein fremdes Netzwerk, so muss er von diesem Netzwerk zusätzlich zu seiner eigenen Adresse (Home Address) eine Adresse – die so genannte COA, Care-Of-Address – zugewiesen bekommen, um dort erreichbar zu sein. Diese Adresse muss er seinem Home Agent mitteilen, damit dieser wiederum die für den Host bestimmten Datenpakete (identifizierbar anhand der Home Address des Host) an diese Adresse weiterleiten kann. Idealerweise soll sich ein Host auch immer wieder von einem Netzwerk abmelden und dies einem Home Agent mitteilen, was in der Praxis allerdings meist nicht geschieht. Kehrt ein Host wieder in sein eigenes Heimnetzwerk zurück, dann meldet er sich bei seinem Home Agent ab weil er nun ja wieder direkt unter seiner Adresse erreichbar ist, damit für ihn bestimmte Pakete direkt zu ihm gelangen und nicht mehr den Umweg über den Home Agent machen.

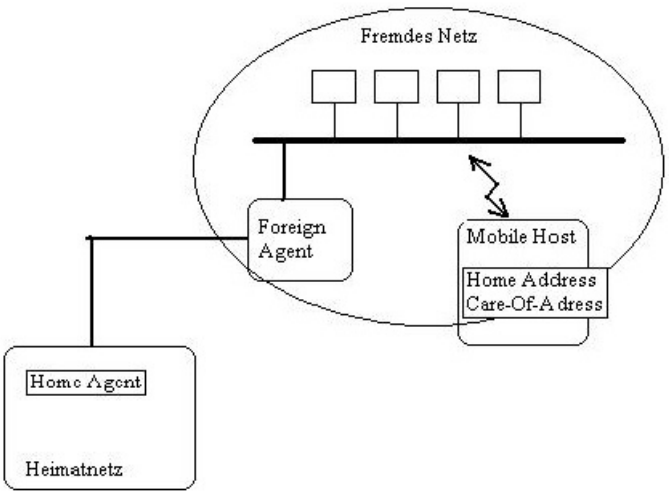


Abbildung 10.11: Mobile IP Registrierung [11]

Zur Vergabe der COA gibt es zwei verschiedene Ansätze: Entweder bezieht der Host selbst die Adresse aus einem Adressraumn (Co Located COA) oder über einen *Foreign Agent*, wie in Abbildung 10.11 von [11] dargestellt. Bei diesem Foreign Agent meldet sich der

Host erst an und bekommt im erfolgreichen Fall eine Adresse aus diesem Netzwerk. Im Falle der Existenz eines Foreign Agents dient dieser auch als Zugangspunkt ins Netzwerk. Der Home Agent leitet für den Host bestimmte Meldungen an den Foreign Agent weiter, welcher sie dann dem Host ausliefert. Der Host selbst ist nicht gezwungen seine Antworten ebenfalls über seinen Home Agent zu schicken. Für manche Anwendungen ist dieser Weg – obwohl er indirekt und damit länger ist – dennoch empfehlenswert, da damit je nach Implementierung Probleme mit Firewalls entfallen können.

### Registrationsprozess mit der *Mobile IP*-Erweiterung für *Diameter*

Mittels diesen Mechanismen von *Mobile IP* ist es also möglich das Netzwerk zu wechseln und weiterhin erreichbar zu bleiben. Wir sind bisher davon ausgegangen, dass eine Registrierung bei einem Foreign Agent möglich ist. Für solche Aufgaben wurde *Diameter* auch geschaffen. Wir wollen nun genauer betrachten, wie eine Registration über die Kombination von *Diameter* und *Mobile IP* stattfindet. In Abbildung 10.12 ist ein Registrationsprozess mittels *Diameter* und *Mobile IP* dargestellt. Betrachten wir erst die einzelnen Komponenten und was bei ihnen geschieht:

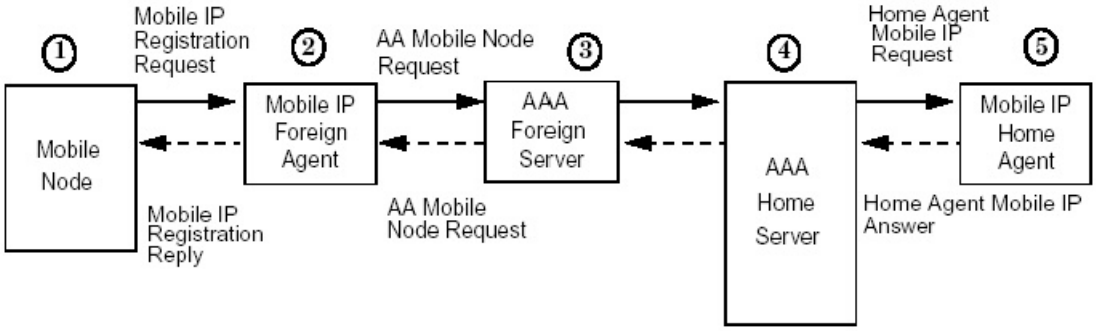


Abbildung 10.12: Registration [12]

1. Mobile Node: Nachdem ein Host (*Mobile Node*) im Bereich eines fremden Netzes ist und sich dort anmelden will, sendet er einen Registrierungsantrag an den *Foreign Agent*

2. Mobile IP Foreign Agent: Der *Foreign Agent* entpackt den AAA-Teil und sendet ihn zur Validierung an seinen AAA-Server, den *AAA Foreign Server*

3. AAA Foreign Server: Der *AAA Foreign Server* ermittelt anhand der Meldung den jeweiligen *AAA Home Server* des *Mobile Nodes* und leitet dorthin den AAA-Teil zur Überprüfung weiter

4. AAA Home Server: Nach der Validierung des AAA-Teils sendet der *AAA Home Server* einen Routingantrag für den *Mobile Node* (damit die neuen für den *Mobile Node* bestimmte Meldungen nun zum *Foreign Agent* geschickt werden) zum *Home Agent*

5. Mobile IP Home Agent: Der *Home Agent* nimmt die neue Adresse für den *Mobile Node* in seine Routingtabelle auf und sendet die Bestätigungsinformationen auf dem gleichen Weg zurück

#### Netzwerktopologie für *Diameter Mobile IP*

Wie dies alles nun konkret aussieht ist Abbildung 10.13 zu entnehmen. Ich möchte an dieser Stelle darauf hinweisen, dass der oben beschriebene Registrationsprozess einen *Mobile IP Agenten* **und** einen *AAA Server* im **gleichen Netzwerk** voraussetzt.

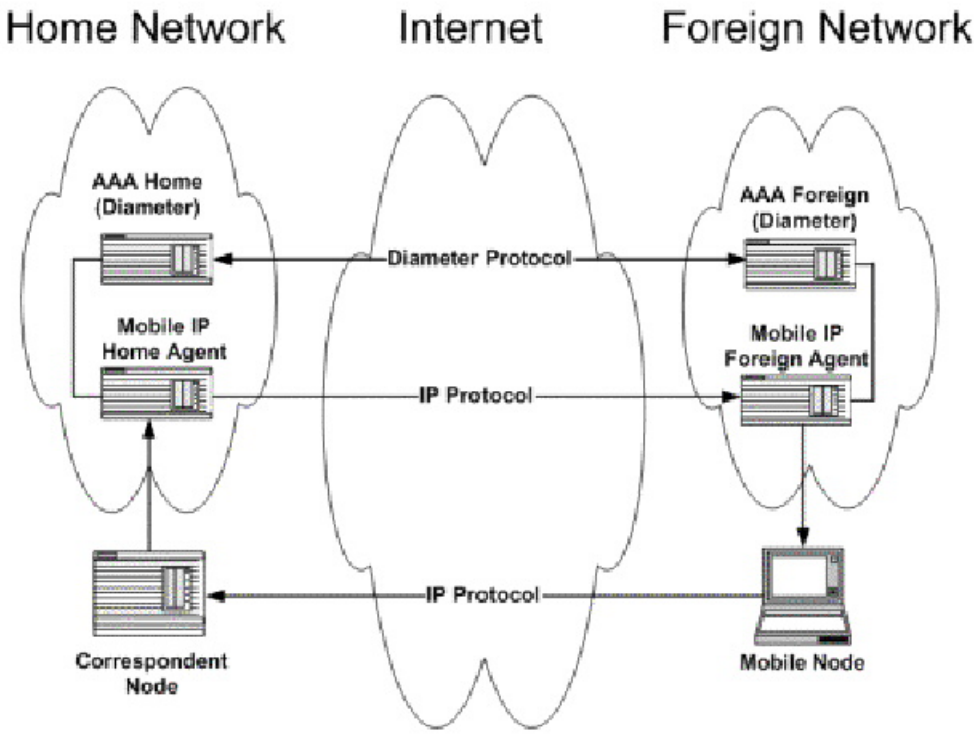


Abbildung 10.13: Topologie [12]

## 10.5 Zusammenfassung

Will *Diameter* die Nachfolge von RADIUS antreten können, muss es auch entsprechende Funktionalitäten zur Authentifizierung (Authentication) der potentiellen Benutzer, deren Zugriffsüberprüfung (Autorisation) und Massnahmen zur verlässlicher Erfassung und Übermittlung von Abrechnungsdaten (Accounting) verfügen – sprich sämtlichen an ein AAA-System gestellten Anforderungen genügen. Gerade das Sicherstellen einer zuverlässigen Übermittlung der Accountingdaten sowie die Wahrung der Konsistenz ist nicht trivial. Dies wird von *Diameter* durch die Nutzung bestehender Protokolle zur verlässlichen Übermittlung (wie bspw. TCP oder SCTP) und durch die Implementierung eigener Massnahmen (Bestätigungsmeldungen, Fehlerverhalten, Timestamps) gelöst.

Trotz der weiten Verbreitung und der grossen Akzeptanz von RADIUS, weist das Protokoll markante Schwächen auf, die früher oder später zu Problemen führen könnten. In Punkto Sicherheit, Flexibilität und Skalierbarkeit wird das Protokoll den heutigen Kommunikationsanforderungen nicht mehr gerecht.

Deshalb wurde schon bald mit der Entwicklung eines neuen Protokolls begonnen. Ziel war es in erster Linie, die Sicherheitslücken von RADIUS zu beseitigen. Andererseits wollte man die Kompatibilität zu RADIUS bewahren, um bestehende Systeme weiterhin nutzen zu können und vor allem, um allfälligen Kosten für die Umrüstung der Infrastruktur so weit wie möglich zu minimieren.

Mit *Diameter* hatte man also eine erste Lösung um den oben genannten Problemen entgegen zu wirken. Das Diameter Basis Protokoll ist im RFC 3588 beschrieben. Das Basis-Protokoll ist für die AAA-Funktionaltitäten zuständig. Für anwendungsspezifische Erweiterungen muss man auf die Applications zurückgreifen, die das Basisprotokoll optional ergänzen. So wird zum Beispiel mittels der NASREQ- oder der Mobile IP-Application die Grundfunktionalitäten um weitere Funktionalitäten ergänzt, um den Zugang zu Network Access Servern zu ermöglich, respektive um mobile Kommunikation zu unterstützen.

Von Diameter erhofft man sich vor allem eine Verbesserung der Sicherheit. Hier wurden ausgeklügelte Fehlererkennungs- und Fehlervermeidungsmechanismen entwickelt, womit man proaktiv auf Ausfälle oder Überlast im Netz reagieren kann. Die Daten müssen mit IPsec oder über TLS verschickt werden. Zusätzlich können die Nutzdaten verschlüsselt werden, damit auch eine Ende-zu-Ende-Sicherheit gewährleistet ist.

# Literaturverzeichnis


[1] Pat R. Calhoun, Comparison of DIAMETER Against AAA Network Access Requirements, draft-calhoun-aaa-diameter-comp-00.txt, April 2000.

[2] B. Aboba, D. Lindyard, „The Accounting Data Interchange Format“, draft-ietf-roamops-actng-07.txt, April 2000.

[3] C. Rigney, S. Willens, A. Rubens, W. Simpson, „Remote Authentication Dial In User Service (RADIUS)“, RFC 2865, Juni 2000.

[4] C.Finseth et al., „An Access Control Protocol, Sometimes Called TACACS“, RFC 1492, Juli 1993.

[5] J. Kohl, C. Neuman, „The Kerberos Network Authentication Service (V5)“, RFC 1510, September 1993.

[6] P. Calhoun, W. Bulley, S. Farrell, „Diameter Strong Security Extension“, draft-calhoun-diameter-strong-crypto-03.txt, April 2000.

[7] R. Stewart et al., „Simple Control Transmission Protocol“, RFC 2960, Oktober 2000.

[8] DARPA(Defense Advanced Research Projects Agency), „Transmission Control Protocol“, RFC 793, September 1981.

[9] P. Calhoun, W. Bulley, „Diameter NASREQ Extension“, draft-calhoun-diameter-nasreq-17.txt“, Juli 2004.

[10] P. Calhoun, C. Perkins, „Diameter Mobile IP Extension“, draft-calhoun-diameter-mobileip-20.txt, August 2004.

[11] „Mobile IP“, http://de.wikipedia.org/wiki/Mobile_IP, Juni 2005.

[12] Hewlett Packard, „Registrationsprozess“, http://docs.hp.com/en/T1428-90053/T1428-90053.pdf, April 2004.

[13] R. Ekstein, Y. T'Joens, B. Sales, O. Paridaens, „Comparison between RADIUS, DIAMETER and COPS“, IETF-§Draft draft-ekstein-aaa-protcomp-01.txt, April 2000.

[14] P. Calhoun, J. Loughney, E. Guttman, G. Zorn, J. Arkko, „Diameter Base Protocol“, RFC 3588, September 2003.

[15] „Open Source Projekt Diameter“, Homepage: *http://sourceforge.net/projects/diameter*

[16] Pat R. Calhoun, Stephen Farell, William, „Diameter CMS Security Application„ , Internet-Draft, draft-ietf-aaa-diameter-cms-sec-04.txt, März 2002.

# Kapitel 11

# Ansätze zur Umsetzung mobiler Ticketing-Systeme

*Thomas Witt und Dani Eichhorn*

*Diese Arbeit gibt einen Überblick der angewandten bzw. sich abzeichnenden Verfahren und Technologien zur Umsetzung von mobilen Ticketing-Systemen. Dabei wird die wirtschaftliche Bedeutung des Marktes mit mobilen Tickets heute und morgen, die Schlüssel-Anbieter auf Seiten Technologie, Veranstalter und Abrechnungspartner, sowie deren Zielpublikum beleuchtet. Anhand eines erarbeiteten Kriterienkataloges zur Klassifizierung von Technologien für mobile Ticketing-Systeme werden existierende und sich in Ausarbeitung befindliche Verfahren aus einem technischen Blickwinkel bewertet.*

## Inhaltsverzeichnis

## 11.1 Einführung ins Mobile Ticketing

Es begleitet uns überall hin, speichert unsere wichtigen Kontakte und ermöglicht uns immer und überall erreichbar zu sein: das Mobiltelefon. Längst haben wir uns daran gewöhnt, täglich Probleme z.B. auf dem Weg zur Arbeit mit seiner Hilfe zu lösen: wir verabreden uns über SMS zum Essen, informieren uns mit dem Handy über die beste Zugsverbindung oder erledigen sogar Bankgeschäfte von diesem Kleingerät aus, das ständig in unserer Nähe ist. Dieses Gerät, von welcher Marke und über welchen Telefonie-Anbieter wir auch immer Anschluss an die Welt erlangen, hat in kurzer Zeit unser Vertrauen gewonnen und wird damit immer mehr zu einem ?Personal Trusted Device?. Dieser vertrauenswürdige Endpunkt des Webs in unserer Hosentasche bietet noch viel Entwicklungspotential für neue Geschäftsmodelle, die mit seiner Hilfe erschlossen werden wollen. Ein erfolgreiches Anwendungsgebiet des Mobiltelefons könnte in Zukunft das Mobile Ticketing werden. Mobile Ticketing bezeichnet das Bestellen, Verrechnen und Verwalten von virtuellen Tickets mit Hilfe von mobilen Geräten. Bei geschätzten 120 Mio. Tickets, die jährlich in der Schweiz zu einem Marktwert von 2 Mrd CHF [27] verkauft werden, ist das Marktpotential für Mobile Ticketing enorm. Dementsprechend gibt es auch bereits verschiedene Anbieter, die mit ihren Lösungen ein möglichst grosses Stück von diesem Kuchen abschneiden wollen. Obwohl die verschiedenen Lösungen vermutlich gleichzeitig auf demselben Endgerät eingesetzt werden können, wäre aus Kundensicht eine standardisierte Abwicklung wünschenswert.

Mobiles Ticketing birgt neben neuen Möglichkeiten aber auch Probleme, die es zu lösen gilt. Insbesondere Sicherheitsaspekte bei der Übertragung und Speicherung der Daten müssen den Anforderungen gerecht werden. Damit mobile Tickets schliesslich von den KonsumentInnen akzeptiert werden, müssen die verwendeten Verfahren und Systeme einen gewissen Grad an Benutzerfreundlichkeit aufweisen. Diese und weitere Kriterien werden in einem späteren Teil anhand eines Kriterienkataloges definiert, der wiederum Grundlage für die Bewertung bestehender und sich in Ausarbeitung befindlichen Verfahren bildet.

## 11.2 Ein mögliches Anwendungsszenario

Heinrich Muster sitzt am Abend nach der Arbeit im Tram, als ihm gerade einfällt, dass er seiner Frau versprochen hat, für den Abend Tickets fürs Kino zu reservieren. Den ganzen Tag während der Arbeit war er unter Druck und konnte sich nicht darum kümmern. Heinrich zieht sein Handy aus der Manteltasche und sucht sich über WAP einen geeigneten Film aus, für welchen er anschliessend Tickets bestellt. Dabei hat er natürlich seinen Namen angegeben und auch den Filmtitel, den Namen der Kinofiliale, die Zeit der Vorstellung und die Anzahl Tickets nicht vergessen. Auf dem Server der Kinokette Metrocinema werden die Plätze reserviert und die eventspezifischen Daten (Datum des Anlasses, Titel), sowie der Name des Besuchers zusammen mit einem zweidimensionalen Barcode zu einer Bildmitteilung zusammengefügt. Der 2D-Barcode enthält die wichtigsten Daten in verschlüsselter Form und kann später von einem Lesegerät einfach ausgelesen werden. Nun erhält Heinrich Muster eine Bestätigung, dass sein Ticket erstellt worden ist.

Anschliessend bezahlt Heinrich die Kino-Tickets über eine der verschiedenen Varianten und erhält kurz darauf das Ticket auf sein Mobiltelefon als MMS. Wenige Stunden später wartet Heinrich mit seiner Frau auf die Öffnung des Kinosaals. Beim Eingang holt er sein Handy hervor, auf dem er das Ticket bereits in die Ansicht geholt hat. Der Kontrolleur am Eingang liest mit seinem Barcode-Lesegerät die Daten vom Mobiltelefon. Das Lesegerät weiss über alle vergebenen Tickets für den bevorstehenden Film bescheid und zeigt dem Prüfer an, dass mit dem Ticket alles in Ordnung ist und es bis jetzt noch nicht verwendet worden ist. Das mit einem speziellen Verfahren strukturierte Bild stellt sicher, dass auch nur der Person, die das Ticket bezahlt hat, Zutritt zum Film gewährt wird.

Dieses Beispiel zeigt sehr schön die verschiedenen Vorteile auf, die mit einem Mobile Ticketing System für Endkunden, Veranstaltungsplattform und den Veranstalter selbst realisiert werden können: Heinrich Muster als Vertreter der Endkunden bekommt eine Möglichkeit mehr, wie er Tickets kaufen kann. Da er beide Eintrittskarten fürs Kino bereits über seine Telefonrechnung bezahlt hat, muss er auch nicht eine halbe Stunde vor Filmbeginn die Tickets abholen, wie er dies früher tun musste. Der Kino-Betreiber kann fortan auf diese Regelung verzichten, er muss ja nun nicht mehr befürchten, auf den Tickets sitzen zu bleiben. Da Heinrich Muster den ganzen Tag über bei seiner Arbeit viel zu tun gehabt hat, wäre es ihm auch nur schwer möglich gewesen, die Tickets über das Internet zu kaufen. Er hat da zwar einen grösseren Bildschirm und Texteingaben (also z.B. die Suche nach den Filmen) sind bequemer zu erledigen. Dank dem Mobile Ticketing kann er aber die tote Zeit im Tram hervorragend dazu nutzen.

Der Kino-Betreiber Metrocinema auf der anderen Seite kann mit steigendem Anteil des Absatzes über Mobile Ticketing die Grösse seines Call Centers reduzieren, was sich stark zu Gunsten seiner Kosten auswirkt. Er kommt zudem näher an den Endkunden heran und macht seine Dienstleistung durch die erweiterten Einkaufsmöglichkeiten attraktiver. Metrocinema zahlt zwar pro verkauftes Ticket eine Gebühr an den Technologie-Anbieter, dafür entfallen aber bei elektronischen Tickets auch die Druckkosten und die Verwendung des mit fälschungssicheren Merkmalen versehenen Ticket-Papiers. Hat sich das System erst einmal etabliert und hat grossen Zuspruch bei den Kinobesuchern gefunden, plant Metrocinema den neuen Vertriebskanal auch von der Marketingseite her stärker zu nutzen. Z.B. könnte registrierten Benutzern auf Wunsch spezielle Angebote auf ihr Handy geschickt werden. Hat z.B. ein männlicher Kunde in der Vergangenheit häufig Action-Filme und gelegentlich Romantische Komödien besucht, wird sein Werbe-Angebot darauf abgestimmt. Von Zeit zu Zeit schickt man ihm auch Sonderangebote für schlecht ausgelastete Kinotage wie den Montag. Um aber die Akzeptanz beim Zielpublikum noch weiter zu erhöhen, gibt Metrocinema einen Teil seiner Mobile Ticketing Kostenvorteile an die Endkunden weiter und macht über das Mobiltelefon gekaufte Tickets billiger als ihre materiellen Pendants.

## 11.3 Markt für Mobile Ticketing-Systeme

In diesem Abschnitt wird der Markt für mobile Ticketing-Systeme grob durchleuchtet. Dabei wird auf die Wirtschaftliche Bedeutung heute und morgen, sowie die Schlüssel-Anbieter und deren strategische Ausrichtung und Zielpublikum eingegangen.

### 11.3.1 Wirtschaftliche Bedeutung heute und morgen

Mobile Ticketing öffnet den Anbietern von Ticketing-Lösungen und deren Zielpublikum die Tore zu einem äusserst lukrativen Markt. Einerseits weichen die Kosten des ?Fulfillments? bei Papier-Tickets den bescheidenen Kosten einer elektronischen Lösung, andererseits profitieren die Ticketkäufer von tieferen Ticketpreisen, erweiterten Informationsmöglichkeiten und schnelleren Prozessen. LogicaCMG [37], ein Pionier-Unternehmen im Bereich Mobile Ticketing, geht von 20 Prozent tieferen Kosten aus [17]. Die MATRIX Solutions GmbH [38] lockt gar mit einer Reduzierung der Vorverkaufs-Gebühren von 60 Prozent [26]. Die Sparte des Mobile Ticketing hat gemäss Informa Telecoms & Media [41] das grösste Potential innerhalb des Mobile Commerce Bereichs (vgl. Abbildung 11.1 [1]).

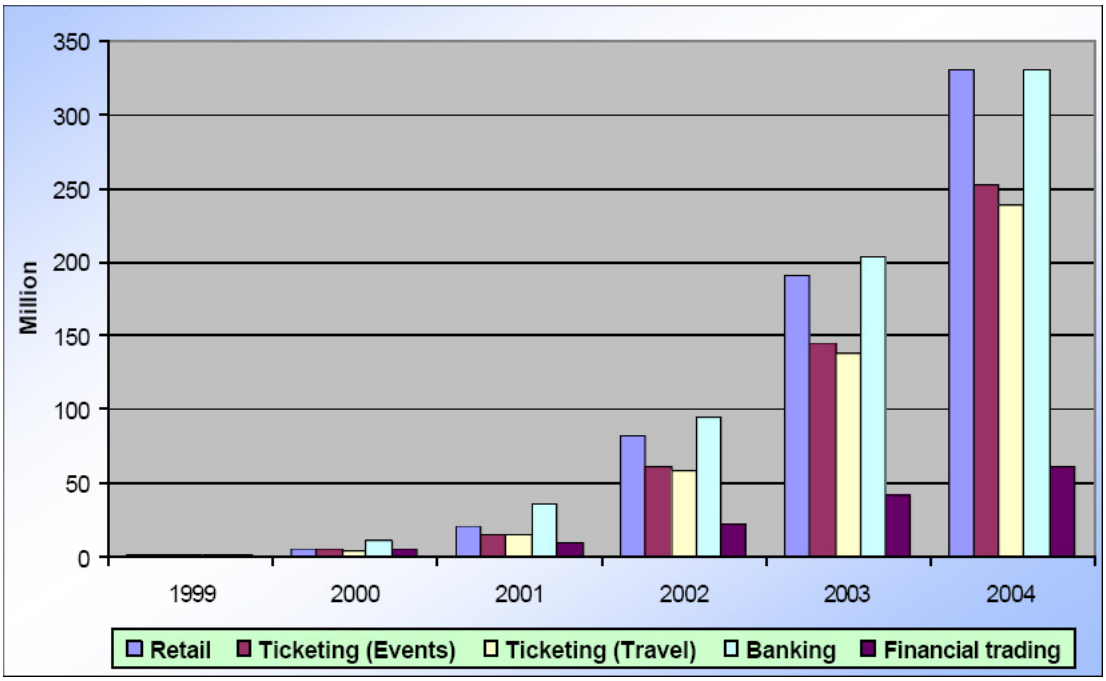


Abbildung 11.1: Wachstum des Mobile Ticketing Marktes [1]

In Westeuropa erwarten die Analysten ein Marktwachstum von $0.5 Billionen im Jahr 2002 auf $2.6 Billionen im Jahr 2007 (LogicaCMG, Ovum [39]). Die Transport- bzw. Reiseindustrie und insbesondere Billiganbieter sind die treibende Kraft hinter diesem Wachstum. Dazu kommt die Tatsache, dass sich Mobiltelefone zu persönlichen, vertrauenswürdigen und weit verbreiteten Geräten entwickelt haben.

Mobile Ticketlösungen wirken sich in mehreren Punkten auf beteiligte Organisationen aus. Die Verfügbarkeit von Angeboten steigt durch mobile Geräte genauso wie die Möglichkeit sachbezogene Kundendaten über entsprechende Kanäle zu liefern. Ausserdem können Angebote aktuell und einfach abgerufen, sowie effizient in betriebswirtschaftliche Prozesse eingebunden werden. Die Mobilität und die dahinterliegende Technologie erlauben ohne grossen Personal- und Materialaufwand Ticketbestellung, -Auslieferung und -Prüfung zu jeder Zeit und an jedem Ort. Die Ticketanbieter (Reise-, Unterhaltungs-, Freizeit- und Transportindustrie) können also durch kundenfreundlichen Service und Kostenersparnisse von Mobile Ticketing profitieren und stellen deshalb die entscheidenden Akteure im

Rennen um den grössten Profit. Schlussendlich wird die Akzeptanz von Mobile Ticketing in der Bevölkerung die Weichen für ein erfolgreiches Wachstum des Marktes stellen. Aus diesem Grund wurde das Pilotprojekt des holländischen Bahnunternehmens NoordNed Personenvervoer BV [40] nicht nur von Technologieanbietern, sondern auch von der Vrije Universiteit Amsterdam [42] begleitet, die neben der Akzeptanz auch das Kauf- und Nutzungsverhalten der Fahrgäste analysiert hat.

Reise- und Transportunternehmen gehören zu den ?First Movern?, die bereits funktionierende und akzeptierte Lösungen einsetzen. Sie bedienen sich jeweils unterschiedlicher Entwicklungen von Technologieunternehmen, die sich dem mobilen Markt zugewandt haben. Veranstalter der Unterhaltungs- und Freizeitindustrie setzen meistens auf Servicepartner, die ihnen neben Soft- und Hardwarelösungen auch Dienstleistungen wie Beratung der Ticketkunden, Koordination, Abrechnung etc. zur Verfügung stellen.

### 11.3.2 Anbieter, strategische Ausrichtung und Zielpublikum

Im folgenden Abschnitt werden die Schlüssel-Anbieter, deren strategische Ausrichtung, sowie deren Zielpublikum mit Konzentration auf den schweizerischen Markt vorgestellt.

#### Technologieanbieter

Technologieanbieter wenden sich hauptsächlich an Business-Kunden, die wiederum das Management der entwickelten Lösungen bis zu den KonsumentInnen übernehmen. Die folgende Aufzählung zeigt die wichtigsten Unternehmen und deren Konzept:

- Die Matrix Solutions GmbH hat sich zum Ziel gemacht, ein schnelleres und preiswerteres Ticketing-System zu entwickeln. Sie verwendet dazu den Data-Matrix-Code, ein zweidimensionaler grafischer Code mit relativ hoher Speicherkapazität, der seit Ende der 80er Jahren vorwiegend bei Produktionsprozessen in der Pharmazie und Automobilindustrie im Industrie-Einsatz ist. Das Verfahren ist patentiert und unter anderem bei LogicaCMG im Einsatz. [26]
- LogicaCMG ist ein weltweit tätiges Unternehmen in den Bereichen Management und IT Beratung, sowie Systementwicklung und -integration. Die für NoordNed entwickelte mobile Ticketlösung macht den Erwerb von Zugtickets für die Kunden einfacher und schneller, wobei eine Tür für weitere Dienstleistungen offen bleibt. Das Ziel des Projektes ist nicht die Verbreitung von mobilen Paymentlösungen, sondern die Etablierung der mobilen Tickets. Um diese Etablierung zu erreichen, braucht es eine sichere Lösung, die alle Risiken ausschaltet und damit möglicher Skepsis zuvorkommt. 2003 ist das Projekt von der Testphase in den echten Betrieb gewechselt. Neben den Reisenden sind auch die Mitarbeiter von dem neuen Dienst begeistert. Durch die Zusammenarbeit mit der SKIDATA AG [44], einem Unternehmen der Kudelski Gruppe [43], die sich als weltweit führenden Anbieter von digitalen Sicherheitslösungen begreift, wird auf den Wunsch nach Komfort, Mobilität und grenzenloser Freiheit unserer Gesellschaft eingegangen. [24]

- Siemens Business Services [45] ist ein Tochterunternehmen des grössten Mobiltelefon-Herstellers und Technologieführer im Bereich mobiler Anwendungen. Wie NoordNed geht auch Siemens Business Services davon aus, dass die Zufriedenheit der Endkunden den entscheidenden Erfolgsfaktor ausmacht. Die mobile Ticket-Lösung soll einfache Bedienbarkeit für den Kunden, Fälschungssicherheit sowie eine schnelle und nachvollziehbare Abwicklung der Kundenabrechnung garantieren. ?Der Fahrgast erwartet nicht nur einen vertrauensvollen Umgang mit seinen persönlichen Daten, sondern auch eine hundertprozentige Verfügbarkeit der mobilen Anwendung. Um die Verfügbarkeit der Lösung zu gewährleisten, bieten wir ein ganzheitliches Service-Konzept mit einer zentralen Anlaufstelle an, das alle technischen Anforderungen abdeckt?, erörtert Michael Adler, Practice Manager Security bei Siemens Business Services. [9]

- Die Lösung für den öffentlichen Personen-Nahverkehr der Siemens VDO Automotive AG [46] zeichnet sich durch eine automatische Anwesenheitserfassung aus. Das Motto ?einfach einsteigen und fahren? wird durch die so genannte Raumerfassung und einen patentierten Systemansatz ermöglicht. Das Ziel dieses Ansatzes ist es, Check-in und Check-out Prozesse ohne Aufwand und Wartezeiten für den Fahrgast zu realisieren, ganz nach dem Vorbild der Mobiltelephonie oder der Strom-Nutzungserfassung. Auch hier heisst die Strategie ?der Kunde ist König?. Das System basiert auf einer Karte mit RFID-Transponder der von Weckfeldern beim Einstieg in das Verkehrsmittel aktiviert und während der Fahrt erfasst wird. Die Akzeptanz dieser Technologie ist laut der Siemens VDO Automotive AG insbesondere auch bei älteren Leuten hoch, weil es sich um ein absolut bedienungsfreies System handelt. [36]

- Nokia [47], Sony [49] und Royal Philips Electronics [50] setzen bei Ihrem Pilotprojekt in Zusammenarbeit mit dem Rhein-Main Verkehrsbund auf eine Lösung mit Mobiltelefonen, die mit einer Near Field Communication (NFC) Hülle ausgestattet sind. Sie nutzen den Vorteil der Kompatibilität mit bereits installierten kontaktlosen Infrastrukturen wie diejenige in Hanau, wo das Pilotprojekt stattfindet. Diese grossen Mobiltelefon-Hersteller versuchen also ganz bewusst einen neuen Markt für spezielle Mobiltelefone zu öffnen. [21][22]

- mTicket [51] ist nach eigenen Angaben das erste Unternehmen, das eine mobile Ticketlösung angeboten hat. Die elektronischen Tickets auf mobilen Geräten versprechen insbesondere finanzielle Vorteile und können als verschlüsselte SMS, MMS Bilder oder Barcodes vertrieben werden. [18]

- Die Gavitec AG [52] hat sich auf flexible und einfach zu bedienende Code-Lesegeräte für mobiles Marketing und Ticketing spezialisiert. In Zusammenarbeit mit Vodafone [48], einem Anbieter von Telekommunikations-Dienstleistungen in 28 Ländern, und der CTS-Eventim AG [53], dem führenden Ticketing-Unternehmen Deutschlands, hat die Gavitec AG ihre führende Position ausgebaut. [23]

- RegiSoft [54] entwickelt Software, die es den Partnerunternehmen erlaubt, mobiles Marketing, interaktive Dienstleistungen, personalisierte und mobile Verbreitung von Inhalt, sowie mobiles Ticketing in bereits existierende Systeme zu integrieren. Der so genannte ?World Trade Server? wird als Flaggschiff bezeichnet und ermöglicht

Ticketbestellungen überall und zu jeder Zeit über mobile Geräte. Kabellos, simpel und effektiv sind die Schlagworte, die als strategische Ausrichtung verstanden werden können. [19]

- Die Teltix GmbH [55] hält Mobile Ticketing hinsichtlich Investitionsbedarfs, technologischer Flexibilität und Kundenbindung für den idealen Vertriebsweg für Fahrscheine im Öffentlichen Personennahverkehr und begreift sich als Service Provider für die Abwicklung des Ticketverkaufs über Mobiltelefone. SMS-Tickets, beste Preise, Informationen und Services bilden die Grundsteine des Geschäftsmodells. [11]

- VisionOne [56] ist ein global tätiges Software-Unternehmen aus St. Gallen und bietet eine RFID-Karte an. Das Ziel ist es, Beratung, Entwicklung und Abwicklung der Geschäftsprozesse aus einer Hand anzubieten. [30]

- ELCA [57]ist einer der grössten IT-Integratoren und -Dienstleister der Schweiz und hat eine eigene Lösung entwickelt, die insbesondere auf Sicherheit setzt. Das Produkt Mobile Secutix, ermöglicht die Bestellung und Lieferung von sicheren Tickets auf MMS fähige Mobiltelefone. Das auf MMS basierende Ticketing stellt im Hinblick auf die Verwendung der integrierten Kamera von MMS-Mobiltelefonen eine Weltneuheit dar, die bereits für eine rein visuelle Kontrolle eine hinreichend hohe Kontrollsicherheit bietet. Das mobile Ticket wird aus dem eigentlichen Ticket und einem eingesandten Bild des Nutzers zusammengestellt. [27]

- Bei der schweizerischen Up-Great AG [58] hat sich zum Ziel gemacht, Produkte zu entwickeln, die eine hohe Kundenzufriedenheit erreichen. Die Division Xsmart findet in den Bereichen Ausweiserstellung, Dokumentensicherheit, E-Ticketing und Brand-Protection Einzug. Die Fälschungssicherheit der so genannten Smart E-Tickets steht im Vordergrund. In Zusammenarbeit mit Swisscom Solutions [59] und der Tic Tec AG [60] ist Verfahren entstanden, das sich auf Veranstaltungen im Kultur- und Freizeitbereich konzentriert. [28]

- Auch die Eidgenössischen Technischen Hochschulen in Zürich [61] und Lausanne [62] haben Lösungen für mobiles Ticketing entwickelt. Eine davon ist z.B. bei der Cinerent Open Air AG im Einsatz. Die Tickets können sowohl auf mobile Geräte geladen, als auch zuhause ausgedruckt werden.

### Veranstalter und Abrechnungspartner

Die folgende Aufzählung zeigt die grössten Event-Veranstalter der Schweiz und beschreibt deren Tätigkeitsfeld:

- Die Good News AG [63] ist die grösste Veranstalterin der Schweiz. 51 Prozent des Aktienkapitals befindet sich im Besitz der Deutschen Entertainment AG (DEAG [64]), 44 weitere Prozent gehören der Ringier AG [65] und André Béchir, einer der Gründer, hält die restlichen fünf Prozent. Die Good News AG veranstaltet grosse Rock- und Popkonzerte, sowie Opernproduktionen und verschiedene Shows.

- Die Freddy Burger Management Unternehmensgruppe [66] ist Veranstalterin im Bereich Tanz, Theater, Musik und Sport. Der Tätigkeitsbereich deckt auch Artist Management, Gastro Management, Projekt-Planung und -Management, Finanz- und Treuhand-Dienstleistungen, Rechtsberatung, sowie Public Relations und TV-Produktionen ab.

- Live Music Production [67] veranstaltet grosse Popkonzerte, Kongresse und Shows in der Westschweiz.

- Der Schweizerische Fussballverband (SFV [68]), wohl die bedeutendste Sportorganisation der Schweiz, zählt heute in 1'500 Vereinen 11'200 Mannschaften und 208'000 Aktivspieler. Nach der Gründung im Jahr 1895 gehörte der SFV zu den sieben Ländern, welche 1904 den Weltfussballverbad FIFA ins Leben riefen.

- Die Schweizer National-Circus AG (Zirkus Knie [69]) ist die grösste Zirkus-Veranstalterin der Schweiz. Auf der Tournee werden 200 Mitarbeiter aus 16 Nationen beschäftigt, die von rund 150 Tieren begleitet werden.

- Opus One [70] ist ein Gemeinschaftsprojekt des Paléo Festival Nyon [71] und der Montreux Jazz Festival Foundation [72]. Das Tätigkeitsfeld umfasst Rock-, Pop-, Jazz- und Blueskonzerte, sowie diverse Shows.

- Die All Blues Konzert GmbH [73] machte sich vor allem mit der Durchführung hochklassiger Jazzkonzerte in den grossen klassischen Konzertsälen der Schweiz einen guten Namen.

- Die Free & Virgin Group [74] ist neben der Good News AG die zweite grosse Veranstalterin von Rock- und Popkonzerten in der Deutschschweiz. Daneben werden auch Jazz- und Klassikkonzerte organisiert.

Im Reise- und Transportbereich ist die SBB AG (Schweizerische Bundesbahnen) [75] die grosse Ticket-Anbieterin. Sie beschränkt sich allerdings nicht auf den erwähnten Bereich, sondern versucht auch im Event-Bereich Fuss zu fassen.

Weil sich der Markt des Mobile Ticketings insbesondere auf Mobiltelefone ausrichtet, gehören selbstverständlich alle Telekom-Unternehmen zu den potentiellen Mitspielern. In der Schweiz sind das die Swisscom-Gruppe, sunrise [76] und die Orange AG [77], welche zu France Telecom [78] gehört. Diese Unternehmen können nicht nur durch Partnerschaften zu Anbietern von Mobile Ticketing Mehrwert generieren, sondern besitzen bereits eine eigene Infrastruktur, die sie ohne grossen Aufwand um diese Dienstleistung erweitern können. Auch bei sämtlichen Transport- und Reiseunternehmen ist eine Infrastruktur bereits vorhanden oder wird gemietet. Ausserdem haben diese Unternehmen Erfahrung mit dem Ticketing-Geschäft, was ihnen Vorteile z.B. im Bereich der Kundenbetreuung verschafft. Die schweizerischen Bundesbahnen bieten bereits Tickets, die zuhause ausgedruckt werden können, an, was als Vorstufe zu Mobile Ticketing gesehen werden kann. Im Folgenden werden die wichtigsten Abrechnungs- und Servicepartner mit Fokus auf die Schweiz aufgelistet:

- Ticketmaster [79], ein Tochterunternehmen der Clear Channel Inc., ist das führende Ticketing-Unternehmen der Welt und hat sich zum Ziel gemacht, die besten Systeme und Dienstleistungen für Business-Kunden, sowie komfortablen, sicheren und fairen Zugang zu Tickets für Konsumenten anzubieten. Eine Lösung, die es den Ticketkunden erlaubt, die Tickets zuhause auszudrucken, existiert bereits. Der Schritt zu Tickets auf mobilen Geräten dürfte also auch hier nur eine Frage der Zeit sein.

- Die TicTec AG vertreibt unter anderem mit dem bereits erwähnten Verfahren der Up-Great AG Tickets in der Schweiz mit Konzentration auf den Raum Basel. Innovatives Ticketing soll das Unternehmen konkurrenzfähig machen.

- Die Cinerent Open Air AG [80] betreibt eine Ticketvertriebsorganisation in der Schweiz, die auf möglichst tiefe Preise setzt, und ist gleichzeitig auch Veranstalter von Unterhaltungs-Veranstaltungen. Sie wurde bereits im Zusammenhang mit der Entwicklung der ETH Zürich erwähnt.

- Die schweizerische Ticketcorner AG [81] steht kurz vor dem Eintritt in den bereits stark umkämpften Markt mobiler Ticketlösungen. Sie ist weltweit mit Konzentration auf den deutschsprachigen Raum tätig und unterhält ihr eigenes Software-Entwicklungszentrum in Bad Homburg. Sie gehört wie die SKIDATA AG zur Kudelski Gruppe, hat also beste Voraussetzungen sowohl für den Markt der RFID bestückten Smart-Cards, als auch für denjenigen der elektronischen Tickets auf Mobiltelefonen.

## 11.4 Kriterienkatalog Mobile Ticketing

Im Folgenden soll ein Kriterienkatalog zusammengestellt werden, mit dem die verschiedenen existierenden und zukünftigen Lösungen für Mobile Ticketing beurteilt werden können. Zur Erstellung des Kriterienkatalogs wurden Informationen über die verschiedenen sich in Anwendung befindenden Produkte und über die theoretischen Konzepte gesammelt. Aus diesen wurden dann wesentliche Unterschiede aber auch Gemeinsamkeiten herausgearbeitet und zu einem Kriterien-Katalog abstrahiert und zusammengefasst.

### 11.4.1 Unterscheidung von Tickets

Zuerst muss für den Kriterienkatalog definiert werden, was ein Ticket überhaupt ist und welche Formen davon möglich sind:

Ein Ticket ist ein Nachweis für den Zugang resp. für die Benutzung einer bestimmten Dienstleistung. Mobile Ticketing ist dabei eine elektronische Ausprägung dieses Nachweises mit Hilfe von mobilen Endgeräten. [1]

**Virtuelle Tickets**

Bei diesem Typ von mobilen Tickets bleibt der Nachweis auf dem Server. Der Bezug der Dienstleistung kann dann nur über eine sichere Authentifizierung des Benutzers z.B. mit seiner Identitätskarte erfolgen. Airline Tickets, bei denen man am Check-In die Boarding-Karte erhält, wenn man die Referenz-Nummer zusammen mit der Identitätskarte vorweist, sind Beispiele für diesen Tickettyp. Da der Endkunde nie den ganzen Nachweis der Bezugsrechte in seinen Händen hält, wird dieser Typ hier als Virtuelles Ticket bezeichnet.

**PTD Tickets**

Beim anderen Typ befindet sich der Nachweis der Bezugsrechte direkt auf dem mobilen Endgerät. Dies hat gegenüber Virtuellen Tickets den Nachteil, dass viel mehr Aufwand für die Fälschungssicherheit der elektronischen Tickets geleistet werden muss. Dagegen weist diese Form den entscheidenden Vorteil auf, dass keine Online-Verbindung mit dem zentralen Server vorhanden sein muss. Dies ist ein entscheidender Vorteil bei AdHoc-Veranstaltungen ohne gute (Netzwerk-)Infrastruktur oder bei Fahrausweiskontrollen im öffentlichen Verkehr. Da hier der Nachweis auf dem persönlichen, vertrauenswürdigen Endgerät gespeichert wird, wollen wir diese Form als Personal Trusted Device Tickets oder kurz als PTD-Tickets bezeichnen.

**Ein- und Mehrfachverwendung der Tickets**

Beide oben genannten Typen sollten die Möglichkeiten aufweisen, Ein- aber auch Mehrfachbezüge der Dienstleistung zu erlauben. Hier stellt sich die Frage, wie z.B. Monatsabonnements oder Mehrfahrtenkarten für den öffentlichen Verkehr realisiert werden können, die nicht übertragbar sein sollen. Oder wie kann ein Messebesucher das Messegelände verlassen und zu einem späteren Zeitpunkt wieder Zutritt erlangen? Wie kann dabei Missbrauch verhindert werden?

### 11.4.2 Aspekte des mobilen Endgeräts

Je nachdem, wie der Prozess des Ticketerwerbs und der Authentifizierung schlussendlich aussieht und welche Form von Tickets zum Einsatz kommt (virtuell oder PTD), werden unterschiedliche Anforderungen an das Endgerät beim Kunden gestellt.

Bei Einsatz als PTD Tickets sind der Download und die Speicherung eines Objekts auf das Mobiltelefon feste Bestandteile des Konzepts. Dabei nimmt das Endgerät auch gleich eine Funktion als ?Ticketwallet?, also als Brieftasche für die Tickets ein. Die verschiedenen auf dem Mobiltelefon vorhandenen Tickets sollten vom Endbenutzer angeschaut aber auch gelöscht werden können und dies alles mit einer einfach zu bedienenden Benutzerschnittstelle. Dabei dürften vom Mobiltelefon-Hersteller unabhängige Java-Lösungen stark im Vorteil sein. Denn viele Anwendungen von Mobile Ticketing bedienen lokale Märkte (z.B. ÖV). Möglicherweise werden viele verschiedene Systeme zum Einsatz kommen, wobei

die dazu benötigten Applikationen auf dem Handy einfach nachgeladen werden können. Trotzdem wäre natürlich aus Konsumenten-Sicht ein Standard im Mobile Ticketing wünschenswert. Dies könnte die Zuverlässigkeit der Technik und damit auch die Akzeptanz bei den EndbenutzerInnen erhöhen.

### 11.4.3 Sicherheit

Die Sicherheit beim Mobile Ticketing hat einige vielschichtige Aspekte. Denn wie immer bei Transaktionssystemen gibt es immer mindestens zwei Seiten, die sich auf das System verlassen müssen. Kann ein Transaktionspartner gar zu leicht betrogen werden, verliert es schnell an Vertrauen und wird dann nicht über längere Zeit erfolgreich sein. Sicherheit und Verlässlichkeit einer Lösung sind also für alle Teilnehmer an diesem virtuellen Ticketmarkt entscheidende Erfolgsfaktoren und verdienen besondere Aufmerksamkeit.

#### Datenschutz für den Endkunden

Datenschutz bezeichnet Massnahmen und Regelungen, die dem Missbrauch von personenbezogenen Daten entgegenwirken. Z.B. könnte ein Veranstalter an der Eingangskontrolle zu einem Konzert alle bisher gekauften e-Tickets seiner Kunden auslesen. Also auch Tickets von anderen Veranstaltern, die aber über das gleiche System hergestellt worden sind. Dies kann schlussendlich nicht im Interesse des Endkunden sein, da damit leicht ein Profil seiner Vorlieben erstellt werden kann, ohne dass er es wahrnimmt.

#### Bedeutung von Transaktionssicherheit

Vom Moment, an welchem der Endkunde sein Ticket bestellt, bis zu jenem Zeitpunkt, an welchem er die Dienstleistung gegen Vorweisen des elektronischen Tickets bezieht, können einige Schritte schief gehen. Niemand erwartet von einem technischen System dieser Komplexität 100 Prozent Zuverlässigkeit. Zu viele verschiedene Komponenten müssten völlig störungsfrei zusammenarbeiten. Wichtig aber ist, dass beim Auftreten einer Störung keine Schäden entstehen, die eine Seite benachteiligen. Was geschieht, wenn der Endkunde sein Ticket zwar bereits bezahlt hat, das System aber während der Transaktion ausfällt und er nie ein Ticket für sein Geld bekommt? Wie reagiert das System, wenn es ein e-Ticket an ein Handy schickt, das über keinen freien Speicherplatz mehr verfügt? Kein System, welches auf dem Markt Erfolg haben will, darf hier für den Endkunden Schwächen aufweisen. Schon wenige Fehler würden das Mobile Ticketing diskreditieren und der Technologie somit längerfristig grossen Schaden zufügen.

#### Authentifizierung beim Bezug der Dienstleistung

Ein ganz wichtiges Thema in Bezug auf Sicherheit beim Mobile Ticketing ist die Authentifizierung des rechtmässigen Bezügers der Dienstleistung sowie die Authentizität des

Tickets. Herkömmliche Monats- oder Jahresabonnements des öffentlichen Verkehrs, die meist an eine bestimmte Person gebunden sind, verwenden einerseits Fotos der Person und andererseits (fälschungssichere) Merkmale auf dem Ticket. Wie kann nun bei elektronischen Tickets diese Kombination mit hoher Zuverlässigkeit und geringem Kontrollaufwand erreicht werden? Digitale Bilder ohne weitere Sicherheitsmerkmale oder Verschlüsselungen sind einfach zu manipulieren, auch auf mobilen Endgeräten. Billig-Fluglinien verlangen vom Reisenden den Passport oder die Identifikationskarte. Hier wird die Verantwortung zur Erstellung eines sicheren Authentifizierungsdokuments quasi an den Staat übergeben. Im Flughafen-Umfeld kann aber auch vorausgesetzt werden, dass der Reisende so ein Dokument dabei hat. Im öffentlichen Verkehr, wo schnell kontrolliert werden muss, kommt dies nicht in Frage.

Zu beachten ist hier auch, dass nicht alle Tickets personengebunden sind. Meistens sind z.B. Einfach-Fahrkarten nicht an eine Person gebunden, ein weiteres Sicherheitsmerkmal fällt damit weg.

#### Validierung des Tickets

Neben der Authentifizierung des rechtmässigen Rechtebezügers stellt die Validierung des Tickets am Kontrollpunkt eine grosse Herausforderung dar. Denn eine spezielle Eigenschaft von digitalen Gütern ist im Allgemeinen ihre einfache, verlustfreie und kostengünstige Vervielfältigung. TrickbetrügerInnen könnten also vom Anbieter ein Original erwerben, dieses vervielfältigen und diese Kopien mehrfach an gutgläubige KonsumentInnen verkaufen. Bei der Eingangskontrolle zu einem Konzert muss nun aus Sicht des Veranstalters mindestens sichergestellt werden, dass für jedes verkaufte Ticket auch nur ein Zutritt gewährt wird. Weiterhin wäre es aus Konsumentensicht wünschenswert, wenn es für den Endkunden selbst eine einfache Möglichkeit zur Echtheitsüberprüfung des erworbenen Tickets geben würde. So könnte generell der Betrug mit elektronischen Tickets erschwert und damit für die Endkunden die Attraktivität des Mobile Ticketing erhöht werden.

### 11.4.4 Aspekte der Benutzerfreundlichkeit

Viele Anbieter versuchen ihre Mobile Ticketing Lösung mit dem Argument zu verkaufen, dass es für die Konsumenten viel einfacher wird, Tickets zu bekommen. Deshalb sollten sich die Endkunden nicht mit komplizierten Programmen oder kryptischen SMS-Codes herumschlagen müssen. Wünschenswert wäre eine einfache und intuitive Oberfläche, die für Jedermann leicht zu erlernen ist.

Usability beschränkt sich aber nicht nur auf den Endkunden selbst, sondern z.B. auch auf das Handling bei der Kontrolle bzw. Validierung der Tickets. Gerade weil das Mobiltelefon immer mehr zum Trusted Personal Device wird, möchte es der Endkunde nur ungern zur Kontrolle aus der Hand geben. Oder wie viel Infrastruktur ist bei der Kontrolle nötig? So kann z.B. für eine grosse Open-Air-Veranstaltung nicht immer davon ausgegangen werden, dass jeder Kontrolleur eine Verbindung zum zentralen Server hat. Lesegeräte sollten gerade für den öffentlichen Verkehr leicht und handlich sein und auch unter widrigen Umweltbedingungen (z.B. Sonnenlichteinstrahlung) zuverlässig funktionieren.

### 11.4.5 Kosten von Mobile Ticketing Lösungen

Immer wenn eine neue Technologie eine ältere ablösen oder zumindest ergänzen soll spielen die Kosten für Konsumenten, Veranstalter und Technologie-Lieferanten eine entscheidende Rolle. Eine neue Ticketing Lösung sollte also im Grossen und Ganzen höchstens die Kosten verursachen, welche auch herkömmliche Tickets unter dem Strich ausmachen.

#### Kosten für den Konsumenten

Wie immer Lösungen beim Mobile Ticketing aussehen, Daten müssen zur Bestellung oder für das Ticket selbst mittels drahtloser Kommunikation übertragen werden. Diese Datenübertragungen verursachen Kosten. Den Konsumenten interessiert daher, welche Kommunikationskosten zu seinen Lasten gehen und welche vom Anbieter übernommen werden.

#### Kosten für den Veranstalter

Gehen wir von einem Markt mit den folgenden Teilnehmern aus: der (Konzert-) Veranstalter, der Ticket-Anbieter, der mit seiner Plattform den Verkauf der Tickets organisiert, sowie die Konzertbesucher. Mit welchem Mehraufwand muss nun der Veranstalter rechnen, wenn er bei seiner Veranstaltung mobile Tickets anbieten will?

#### Kosten für den Anbieter

Für den Ticket-Anbieter ist bei der Einführung einer mobilen Ticket-Lösung wichtig, mit welchen Kosten er für die Einrichtung und den Betrieb einer Mobile Ticketing Plattform rechnen muss. Dabei werden die einmaligen Infrastrukturkosten vermutlich weniger bedeutend sein, wie die variablen Kosten für den Betrieb. Obwohl der Anbieter mit dieser Möglichkeit die Attraktivität seiner Dienstleistung bei seinen Kunden erhöht, wird für ihn vor allem wichtig sein, ob mit dem Mobile Ticketing Kosteneinsparungen erreicht werden können.

### 11.4.6 Verfügbarkeit von Technologie Standards

Bei der Einführung von neuen Technologien für die Endkunden hat sich immer wieder gezeigt, dass offene Standards sich meist gegenüber proprietären Lösungen durchsetzen können. Standards verringern die Unsicherheit bei Endkunden und das Risiko für Ticket-Verkaufs-Platformen. Denn die kritische Masse für den Erfolg eines Systems wird durch verbindliche Schnittstellen viel schneller erreicht. Bei Netzwerk-Gütern, zu denen man Mobile Tickets auch rechnen kann, setzte sich in der Vergangenheit oft die technisch minderwertige Lösung durch, wenn diese offene Standards definiert hat. Für dieses Marktverhalten haben Oekonomen das Kunstwort Coopetition geschaffen. Wettbewerber in einem Markt (Competitors) sollten in speziellen Situationen zusammenarbeiten (Cooperation), um für alle Seiten Vorteile zu erlangen.

### 11.4.7 Business-Models im Mobile Ticketing Markt

Kein Betreiber wird in eine neue Technologie investieren, wenn sich damit nicht Vorteile für sein Geschäft realisieren lassen. Für Event-Tickets könnte man sich vorstellen, dass der zentrale Ticket-Anbieter die Technologie mit der Infrastruktur von einem Spezialisten einkauft und danach selbst betreibt. In Anbetracht der schellen Entwicklung und der hohen Komplexität dürfte es sich aber für viele Anbieter nicht lohnen, das System selbst zu betreiben. Viel eher wird die Mobile Ticketing Dienstleistung in einem Gesamtsystem vom Spezialisten geleast und zusätzlich pro Transaktion eine Gebühr bezahlt. Dadurch bezahlt der Ticket-Anbieter zwar mehr pro Ticket, er muss sich dafür aber nicht um die Entwicklung der Technologie und deren Unterhalt kümmern.

Wechseln wir nun die Perspektive vom Technologie-Anbieter zum Ticket-Anbieter: Verlangt er von seinen Kunden eine (monatliche) Basisgebühr für die Benutzung der Mobile-Ticketing-Dienstleistung? Wird gegenüber herkömmlichen Tickets ein Aufpreis erhoben oder sind mobile Tickets sogar günstiger als ihre physischen Gegenstücke?

### 11.4.8 Notwendige Infrastruktur

Wenn eine Mobile-Ticketing-Lösung kaum Ansprüche an zusätzliche Infrastruktur hat, ist dies auf jeden Fall ein Vorteil. Zur Infrastruktur gehören bestimmt die Mobiltelefone der KonsumentInnen und möglicherweise Applikationen, die darauf laufen müssen. Dabei stellt sich die Frage, auf welchem Anteil der verwendeten Mobiltelefone das Mobile Ticketing funktioniert. Muss das Handy MMS-fähig sein, wird eine Java-Umgebung benötigt? Werden spezielle Chips für Digitale Signaturen verwendet, die vielleicht erst in zukünftigen Mobilgeräten vorhanden sein werden?

Die meisten Systeme werden zusätzlich zentrale Server benötigen sowie Lesegeräte am Ort, wo die Tickets validiert werden müssen. Zudem stellt sich die Frage, ob bei der Kontrolle eine Online-Verbindung zum zentralen Server vorliegen muss. Dies wäre sehr gut möglich, denn auch bei Fussballstadien mit mehreren Eingängen muss sichergestellt werden, dass jeweils pro verkauftes Ticket einmal Zutritt gewährt wird und so illegale Kopien zumindest dem Anbieter nicht schaden können.

### 11.4.9 Möglichkeiten des Billing

Seit längerer Zeit versuchen Mobilfunkanbieter Zahlungssysteme über das Mobiltelefon zu etablieren. So richtig durchgesetzt hat sich bis jetzt aber noch keines. Würden die Telekom-Unternehmen mit diesem Vorhaben allerdings Erfolg haben, könnten sie mit einem Schlag zu einer akzeptablen Konkurrenz anderer bargeldloser Bezahlsysteme wie Kredit- oder Bankkarten werden.

Diese Entwicklung würde auch dem Mobile Ticketing als Business helfen voranzukommen. Eingesetzt werden heute für höhere Beträge oft sogennante Premium-Service-Nummern. Der Konsument ruft eine bestimmte Nummer an. Dieser Anruf wird ihm entweder auf der

nächsten Rechnung belastet oder von seinem Prepaid-Konto abgezogen. Damit bekommt er dann aber auch Zugang zur gekauften Dienstleistung. Dieses Prinzip funktioniert je nach Telekom-Anbieter besser oder schlechter, oder kann auch auf SMS angewandt werden. Wichtig dabei ist, dass der Mobiltelefonie-Anbieter als Finanz-Intermediär auftritt, ganz ähnlich einer Kreditkartenunternehmung wie American Express oder Visa. Im Gegensatz dazu sind auch andere Arten der Verrechnung denkbar: Angabe der Kreditkarten-Daten usw.

### 11.4.10 Prozess-Geschwindigkeit für den Endkunden

In eine ähnliche Richtung wie die Usability geht das Kriterium der Prozess-Geschwindigkeit. Wie lange dauert es vom Bestellen des Tickets, bis der Endkunde die digitale Eintrittskarte erhält? Wie lange dauert eine Kontrolle? Muss im Bus bei der Validierung der Fahrkarten erst noch lange auf eine Antwort des Servers gewartet werden? Müssen die Konsumenten die Fahrkarten erst mühsam in einem Untermenu suchen?

### 11.4.11 Rahmenbedingungen

Zusammenfassend für alle vorangehenden Kriterien sollte beurteilt werden, ob und mit welchen Einschränkungen die betrachtete Lösung für einen bestimmten Markt eingesetzt werden kann. Die wichtigsten Märkte dabei sind Reise-, Unterhaltungs-, Freizeit- und Transportindustrie. Ticketing-Anwendungen in diesen Märkten unterliegen jeweils besonderen Einschränkungen oder auch Möglichkeiten. So bestehen mobile Tickets in der Flugindustrie bereits heute nur aus einfachen Nummern. Zur Identifikation wird zusätzlich vom Passagier der Passport verlangt, erst dann kann er einchecken.

## 11.5 Existierende und sich in Ausarbeitung befindliche Verfahren aus einem technischen Blickwinkel

In diesem Abschnitt werden existierende und sich in Ausarbeit befindliche Verfahren von mobilem Ticketing aus einem technischen Blickwinkel betrachtet. Dabei werden die für den Markt interessanten Verfahren, die auf einem RFID-Transponder basieren, am Rande erwähnt, während das Hauptaugenmerk auf Verfahren liegt, welche die Autorisierung der Tickets mittels Verbindung zu einer zentralen Datenbank bzw. Verschlüsselungs- und Darstellungsmethoden auf mobilen Engeräten ermöglichen. Zu diesen Verfahren wird jeweils ein Beispiel genannt und mit dem erarbeiteten Kriterienkatalog bewertet.

### 11.5.1 RFID-Tickets

RFID-Tickets haben den grossen Vorteil, dass sie ohne Aufwand gelesen werden können, sodass es keinerlei menschlicher Interaktion bedarf, weder auf der Seite der Anbieter, noch auf jener der KonsumentInnen. Zu den Nachteilen gehören die Kosten sowohl für die Tickets selber, als auch für das Einrichten von Lese-Systemen.

### 11.5.2 Textbasierte Tickets

Seit September 2002 bietet NoordNed in Zusammenarbeit mit der LogicaCMG-Gruppe ein mobiles Ticketingsystem in Friesland und Groningen an [34]. Die Tickets können entweder auf einer Website oder durch Anrufen einer kostenlosen Telefonnummer bestellt werden. Das Web-Interface basiert auf Java-Beans auf einem Tomcat-Server, der auf eine zentrale Microsoft SQL Server Datenbank zugreift. Das intelligente Spracherkennungssystem Speech Mania von Philips fragt bei einem Anruf die notwendigen Daten ab, worauf das Ticket per SMS an das Mobiltelefon geschickt wird. Voraussetzung zur Teilnahme ist ausser einem Vertrag mit einem holländischen Mobilfunkanbieter der Abschluss eines Direct Debit Agreements, das die Abbuchung der Ticketkosten vom Bankkonto der KonsumentInnen ermöglicht. Die Kontrolle der Tickets erfolgt über einen Autorisierungscode. Die Kontrolleure benutzen einen Pocket PC vom Typ iPAQ 3870, auf welchen vor jedem Kontrollgang alle benötigten Daten geladen werden. Um kurzfristig gekaufte Tickets zu erkennen steht der iPAQ über Bluetooth mit einem Nokia 6310i-Mobiltelefon und dieses über GPRS mit der Zentrale in ständiger Verbindung. Auf der Seite der mobilen Geräte kommt die Datenmanagementlösung der Sybase-Tochter iAnywhere Solutions [82] zum Einsatz. Die Synchronisation der Daten zwischen der Ultralite-Datenbank von SQL Anywhere Studio und der zentraler Datenbank geschieht über einen MobiLink-Server. Das Application Service Provider System von LogicaCMG ermöglicht das Senden und Empfangen von SMS und stellt die Verbindung zu den Mobilfunkanbietern, sowie zwischen Bank und Datenbanken her. Weil in fahrenden Verkehrsmitteln, die z.B. durch Tunnel fahren, keine permanente Verbindung mit der Zentrale möglich ist, muss die Kontrolle auch offline möglich sein. Durch die Kombination von UltraLite und MobiLink ist es möglich, zwischen zwei Synchronisationsvorgängen Informationen auf ein mobiles Endgerät zu laden, zu speichern und dort zu bearbeiten.

Sicherheit spielt bei allen Prozessen eine wichtige Rolle. Das LogicaCMG-System überprüft bei jeder Registrierung die Mobiltelefonnummer mittels einer SMS. Neben der PIN-Abfrage bei der Mobiltelefon-Nutzung fragt auch das Spracherkennungssystem PINs für Login und Ticketerwerb ab. Auf dem Ticket befinden sich Informationen zur Anzahl Personen, Kaufzeit und Strecke, sowie ein Autorisierungscode. Bei der Kontrolle müssen die Fahrgäste die letzten drei Ziffern ihrer Mobiltelefonnummer angeben, die auf dem iPAQ neben dem Autorisierungscode stehen. Schlussendlich können die Tickets durch einen Klick auf dem iPAQ entwertet werden. Diese Lösung erhält also in allen Sicherheitskategorien Bestnoten, ohne dass es eine Verschlüsselung auf dem Mobiltelefon braucht. Der Datenschutz ist gewährleistet, solange sich die beteiligten Organisationen an bestehende Datenschutzgesetze halten. Auch die Transaktionssicherheit scheint keine Lücken aufzuweisen, weil bestehende Sicherheitsstandards bei der Übermittlung der Daten genutzt

werden können. Ab dem Zeitpunkt der Registrierung wird zudem die Authentifizierung durch die PINs gesichert und es findet bei der Kontrolle eine Gültigkeitsprüfung durch den Abgleich mit der Datenbank statt. Ist es allerdings ein Anliegen eines Verkehrsunternehmens nicht übertragbare Tickets zu vertreiben, so kann diese Lösung nicht überzeugen, weil das Ticket ohne grosse Schwierigkeiten weitergereicht werden kann.

Nach Angaben der beteiligten Organisationen war die Zuverlässigkeit der verwendeten Produkte das Schlüsselargument bei der Auswahl. Die Erwartungen wurden auch im Bezug auf die Kundenzufriedenheit bei weitem übertroffen. Weil alle benötigten Kommunikationsmittel (SMS, Telefonie und Website) weit verbreitet bzw. bekannt sind und die Ticketbestellung relativ schnell bzw. ohne Anstehen von statten geht, sowie rund um die Uhr möglich ist, erscheint die Benutzerfreundlichkeit dieses Verfahrens ziemlich gross. Die Angabe der Telefonnummern bei der Kontrolle ist allerdings ein Mehraufwand, der verhindert werden könnte. Einzig für Touristen dürfte es nicht brauchbar sein, weil ein Vertrag mit einem holländischen Mobilfunkbetreiber Voraussetzung für die Benutzung ist.

Die Kosten für die KonsumentInnen sind bei diesem Verfahren auf ein Minimum reduziert. Die SMS-Kosten werden von NoordNed übernommen und in Holland fallen für Buchungsposten von Bankkunden grundsätzlich keine Gebühren an. Durch die rund 20 Prozent Einsparungen beim Ticketverkauf kann NoordNed das gesamte Projekt finanzieren, und auch dessen Entwicklung war sehr günstig, denn sie hat nur zwei Wochen gedauert. Es ist also auch für die Anbieter eine günstige Lösung.

Die Benutzung von Übermittlungs- und Sicherheitsstandards hilft solchen Verfahren sich in bestehende Infrastrukturen einzugliedern. Auch von den KonsumentInnen sind keine besonderen Kenntnisse gefordert, denn nach dem Bestellvorgang gestaltet sich der Ablauf wie sie es sich schon von Papier-Tickets gewohnt sind. Einzig ihre Telefonnummer muss ihnen bekannt sein. Die Abrechnung der gekauften Tickets erfolgt über ein Lastschriftverfahren, was den Aufwand reduziert und ebenfalls keinen neuen Prozess mit sich bringt.

### 11.5.3 Bildbasierte Tickets

Bei bildbasierten Verfahren wird das Ticket z.B. in einen zweidimensionalen Code verschlüsselt und als Bild an bildmitteilungsfähige mobile Geräte übermittelt. Es sind verschiedene Typen von zweidimensionalen Codes im Einsatz. Dabei wird zwischen gestapelten und Matrix-Codes unterschieden. Gestapelte Codes sind aneinander gereihte Barcodes, die eine bestimmte Ausrichtung haben und dementsprechend nur in einer Richtung gelesen werden können [32]. Unabhängig von der Leserichtung sind Matrix Codes, welche die Zuverlässigkeit des Lesevorgangs deutlich erhöhen (vgl. Abbildung 11.2).

Ein bereits standardisiertes und patentiertes Ticket dieser Art bietet die Matrix Solutions GmbH an [26]. Sie arbeitet mit dem Data-Matrix-Code, der seit den 80er Jahren im Industrie-Einsatz ist. Es handelt sich um einen zweidimensionalen grafischen Code mit relativ hoher Speicherkapazität. Bei gleicher Fläche enthält der Data-Matrix-Code eine 100fach höhere Speicherkapazität als ein eindimensionaler Barcode [31].

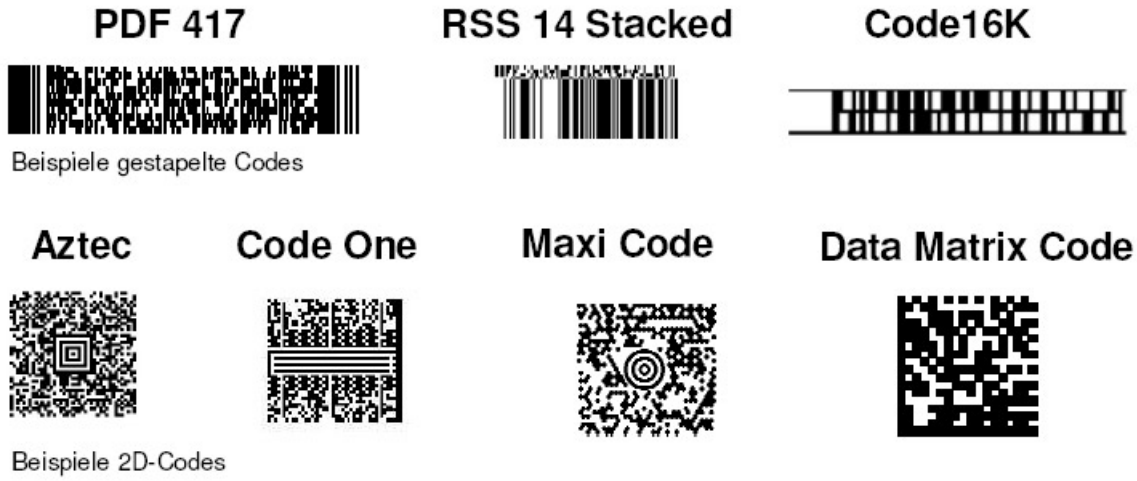


Abbildung 11.2: Zweidimensionale Codes [32]

Das so genannte PicTicket Verfahren kommt unter anderem bei kartenhaus.de [83], einem Vertriebshaus von Tickets im Unterhaltungsbereich, zum Einsatz. Die Veranstaltungen werden mit dem TourOrganizer auf dem Matrix PicTicket DB-Server eingerichtet. Dabei werden alle notwendigen Informationen für die Erstellung der Matrix PicTicket vorgeneriert. Bestellt werden die Tickets im Internet (bzw. je nach Vertragspartner über andere Kanäle wie z.B. das Mobiltelefon bzw. GSM). Dazu müssen die KonsumentInnen aus einer Liste Mobiltelefon-Hersteller und -Typ auswählen, sowie ihre Mobiltelefonnummer angeben. Die Autorisierung findet bereits hier über Benutzernamen und Passwort statt.

Im Inventory-System werden alle kundenspezifischen Daten der Bestellung erfasst, personalisiert und als Datenpaket an den Matrix PicTicket DB-Server geschickt. Dieser generiert ein für das jeweilige Handy spezifisches Matrix PicTicket und schickt es über ein SMS-Gateway an die jeweilige Mobiltelefonnummer. Voraussetzung für den Empfang ist ein bildmitteilungsfähiges Mobiltelefon. Alternativ können die Tickets auch zuhause ausgedruckt werden. Nachdem die KonsumentInnen online bezahlt haben, erhalten sie eine Zahlungsbestätigung und ein eine Bildmitteilung für jedes bestellte Ticket. Auf dieser Bildmitteilung können weitere Informationen (z.B. das Programm bei Festivals, Gutscheine oder Werbung) untergebracht werden. Geht ein Ticket verloren oder wird versehentlich gelöscht, kann es über ein Call Center storniert und gegen eine geringe Gebühr durch ein neues ersetzt werden. Vor der Veranstaltung holt sich der TourManager über eine SSL-Verbindung übers Internet mittels XML-Abfrage alle relevanten Ticket- und Konsumenteninformationen vom Matrix PicTicket Datenbank-Server in seine MySQL-Datenbank ab. Alle Scanner sind über Kabel oder Funk an den TourManager angeschlossen. Die Gültigkeitsprüfung geschieht schnell und sicher mit Hilfe von Hand- oder Drehkreuzscannern am Eingang. Dabei wird der Data-Matrix-Code geprüft und entwertet. Für den Fall, dass ein Matrix PicTicket nicht gültig ist, lässt sich gemeinsam mit dem Kunden am ?TourManager? nachvollziehen, wann und von wem das Ticket gekauft wurde. Zur Auswertung übergibt der TourManager nach der Veranstaltung alle erfassten Daten an den Matrix PicTicket Datanbank-Server (vgl. Abbildung 11.3).

Die Sicherheit dieses Verfahrens ist über vorhandene Sicherheitsstandards gewährleistet.

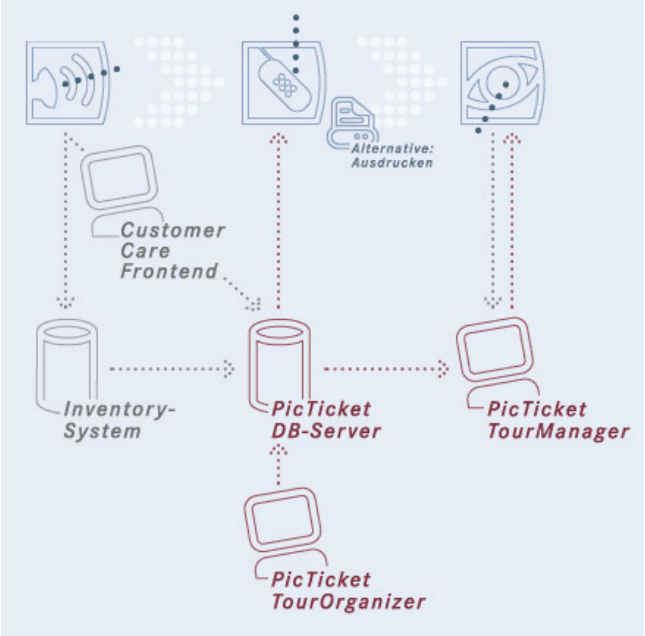


Abbildung 11.3: Verfahren der Matrix Solutions GmbH [26]

Der Datenschutz erscheint besonders gross, weil nur eine bzw. wenige Organisationen am Ablauf beteiligt sind. Die Bestellung bzw. die Transaktion der Daten, sowie die Authentifizierung sind ebenfalls durch standardisierte Verfahren gesichert, zumindest bis zu dem Zeitpunkt, an welchem die Tickets die KonsumentInnen erreichen. Danach sind die KonsumentInnen selber für die Aufbewahrung ihrer Tickets verantwortlich. Auch wenn die Tickets nicht weitergeschickt werden können stellt sich hier das Problem, dass ein Mobiltelefon durchaus den Besitzer wechseln kann, was personenbezogene Tickets mit diesem System alleine nicht möglich macht. Diese Problematik wird von der SBB AG, die zuhause druckbare Tickets vertreibt, durch das zusätzliche Vorweisen eines Ausweises gelöst. Schlussendlich ist die Gültigkeit der Tickets beim Verfahren der Matrix Solutions GmbH durch den Abgleich mit einer Datenbank gewährleistet.

Weil Matrix Solutions und deren Partner eine Gesamtlösung anbieten, ist der Aufwand für die Anbieter der Tickets relativ klein. Zusätzlich erleichtert die Bestellung über Internet die Anbindung an bestehende Systeme und die speziell für den Vertrieb von mobilen Tickets entwickelte Software das Handling des Systems. Umsätze können zeitnah erfasst und zugeordnet werden. Auf der Seite der KonsumentInnen ist kein grosser Unterschied zu Systemen mit Papier-Tickets erkennbar. Eventuell werden weniger Tickets zuhause vergessen, weil die angesprochene Zielgruppe ihre Mobiltelefone stets bei sich tragen dürfte.

Wie bereits erwähnt können durch diese Lösung Vorverkaufskosten von bis zu 60 Prozent eingespart werden. Insbesondere die Möglichkeit, ohne Personalaufwand 24 Stunden am Tag Bestellungen annehmen zu können dürfte die Anbieter und damit auch die KonsumentInnen im Hinblick auf die Kosten überzeugen.

Der verwendete Industriestandard der zweidimensionalen Darstellung ist ein weiterer Vorteil dieses Verfahrens. Der Standard funktioniert zuverlässig und es gibt eine grosse Auswahl an Lesegeräten. Für die Lösung der Matrix Solutions GmbH stehen allerdings Li-

zenzgebühren an, welche aber durch Kosteneinsparungen bei Personal und Fulfillment wieder relativiert werden.

Das Business Modell hat sich als erfolgreich erwiesen. Durch die Zusammenarbeit mit spezialisierten Partnern und die Verwendung von Standards ist diese Lösung sehr flexibel. Mitverdiener sind Softwareentwickler und Hersteller von Code-Lesegeräten, sowie z.B. Veranstalter, die über den zusätzlichen Informationskanal und garantierte Aufmerksamkeit der KonsumentInnen Mehrwertwert generieren können.

Die benötigte Infrastruktur ist im Vergleich zu RFID-Tickets eher klein und wird vom Technologieanbieter zur Verfügung gestellt. Einzig für Veranstaltungsstätten, die so gross sind, dass eine Verbindung der Lesegeräte über Funk- bzw. Kabelverbindungen nicht möglich ist, erscheint diese Lösung unbrauchbar. Hier muss heutzutage eine Verbindung über Satelliten hergestellt werden, was die Kosten erheblich in die Höhe treiben dürfte.

Abgerechnet wird über alle im Internet zur Verfügung stehenden Kanäle, also über Kreditkarte oder Rechnung. Bei Zusammenarbeit mit weiteren Partnern sind aber auch weitere Methoden wie z.B. über die Telefonrechnung bzw. Lastschriftverfahren möglich. Insgesamt umfasst das Einsatzgebiet der bildbasierten Tickets alle Bereiche des Ticketings.

### 11.5.4 Programmbasierte Tickets

Am 16. Februar 2004 startete der Pilotbetrieb eines Projektes des Zweckverbandes öffentlichen Personennahverkehrs Vogtland (ZVV) in Zusammenarbeit mit Siemens Business Services im Ländereck zwischen Tschechien und den drei deutschen Bundesländern Bayern, Thüringen und Sachsen [9]. Die KonsumentInnen haben die Auswahl zwischen zwei verschiedenen Verfahren. Beim ersten Verfahren können sie per Anruf über eine gemeinsam mit dem Fraunhofer-Institut [84] entwickelte Spracherkennungssoftware die gewünschten Tickets bestellen. Beim zweiten Verfahren kommt eine Software zum Einsatz, die zuerst über WAP auf ein Java-fähiges Mobiltelefon geladen werden muss. Dazu müssen sich die KonsumentInnen registrieren, wobei auch eine Bankverbindung angeben kann. Die Bestellung erfolgt dann direkt auf dem Mobiltelefon. Nachdem die Daten an das IT-System des ZVV übermittelt worden sind, wird eine Bestätigung per SMS an das Mobiltelefon zurückgesendet. Die Abrechnung erfolgt über ein Lastschriftverfahren und in Zukunft auch per Telefonrechnung, über Kredit- oder Prepaidkarte. Die im Mobiltelefon gespeicherte Berechtigung wird bei der Kontrolle mit einem Kontrollgerät erfasst, wobei die Berechtigung geprüft und mit der Kontrollnummer abgeglichen wird. Die Kontrolldaten können sofort oder zu einem späteren Zeitpunkt an das Hintergrundsystem übertragen und dort verifiziert werden.

Dieses Verfahren hat im Bereich Sicherheit Vorteile, weil eine speziell für diesen Zweck entwickelte Software auf dem Endgerät eingesetzt wird. Die Übertragung wird durch WTLS (Wireless Transport Layer Security) gesichert. Der Datenschutz ist auch bei dieser Lösung gewährleistet, solange sich die beteiligten Organisationen an die Richtlinien halten. Die Identifizierung der Konsumentinnen wird ebenfalls sichergestellt, solange nur übertragbare Tickets vertrieben werden. Ansonsten braucht es wie bei den anderen vorgestellten Tickets

zusätzlich einen Identitäts-Ausweis. Die Gültigkeitprüfung der Tickets ist besonders sicher, weil zusätzlich zum Datenabgleich mit einem Hintergrundsystem die Berechtigung auf der Java-Applikation gespeichert wird.

Siemens Business Services bietet vollen Service an, was das Entwickeln und Integrieren des Systems anbelangt. Interessierte Unternehmen können sich also zurücklehnen und sich auf die Fertigkeiten der Siemens Mitarbeiter verlassen. Abgesehen vom Laden der Java-Applikation erweist sich dieses Verfahren auch für die KonsumentInnen als sehr benutzerfreundlich, weil die Tickets 24 Stunden am Tag auf der gewohnten Umgebung des Mobiltelefons bestellt werden können.

Über die Kosten für die Entwicklung und Integration des Systems wird in dem von uns konsultiertem Material nicht gesprochen. Um mit anderen Lösungen mithalten zu können, dürfen die Kosten für die Ticket-Anbieter allerdings nicht hoch sein, weil Kosteneinsparung bei allen Lösungen ein Hauptargument darstellt. Für die KonsumentInnen dürften die Kosten höher als z.B. bei der NoordNed-Lösung sein, weil die Java-Applikation über WAP mit dem Hintergrundsystem kommuniziert.

Neben speziellen Kontrollgeräten auf der Seite der Kontrolleure sind auch Java- und WAP-fähige Mobiltelefone Voraussetzung. Das Business Modell setzt also auf eine Eigenentwicklung, die es zwar nicht allen MobiltelefonbesitzerInnen ermöglicht, über dieses System Tickets zu bestellen, bei Erfolg aber weitere Einnahmen für die Entwickler verspricht.

Bis anhin wird zwar nur das Lastschriftverfahren für die Abrechnung eingesetzt, durch die anderen möglichen Verfahren, die in Zukunft noch folgen sollen, erhält diese Lösung in diesem Bereich gute Noten. Insgesamt kann sie für alle möglichen Einsatzgebiete des Ticketings verwendet werden, die übertragbare Tickets erlauben.

### 11.5.5 Neue Ansätze

Neuere Ansätze beschäftigen sich vor allem mit der Verbesserung der Sicherheit der kabellosen Übertragung von elektronischen Tickets, des Kopierschutzes, sowie der Authentifizierung der KonsumentInnen.

Ein besonders erwähnenswertes Verfahren ist dasjenige von Chun-Te Chen and Te-Chung Lu vom Department of Information Management der HUAFAN University in Taiwan [8]. Dieses Verfahren bedient sich dem Verschlüsselungs-Algorithmus Visual Secret Sharing (VSS) und wendet ihn auf K. Fujimura's (NTT LABS) e-Ticket-System an. Das Herzstück bilden zwei Ticketteile bzw. ein Zeitstempel. Bei einem Ticketerwerb wird der erste Teil auf das mobile Gerät übertragen. Der zweite Teil wird erst dann übertragen, wenn der Zeitstempel erreicht wird. Die Prüfung des Tickets ist dann von Auge oder mit einem Scanner möglich (vgl. Abbildung 11.4). Voraussetzung ist eine Java-Applikation, welche das Ticket generiert, und ein bildfähiges mobiles Gerät, das das Ticket empfängt.

Weil vom ersten Teil des Tickets keine Rückschlüsse auf den zweiten Teil gezogen werden kann, verhindert dieses Verfahren das Kopieren des Tickets. Neben der Sicherheit erhöht

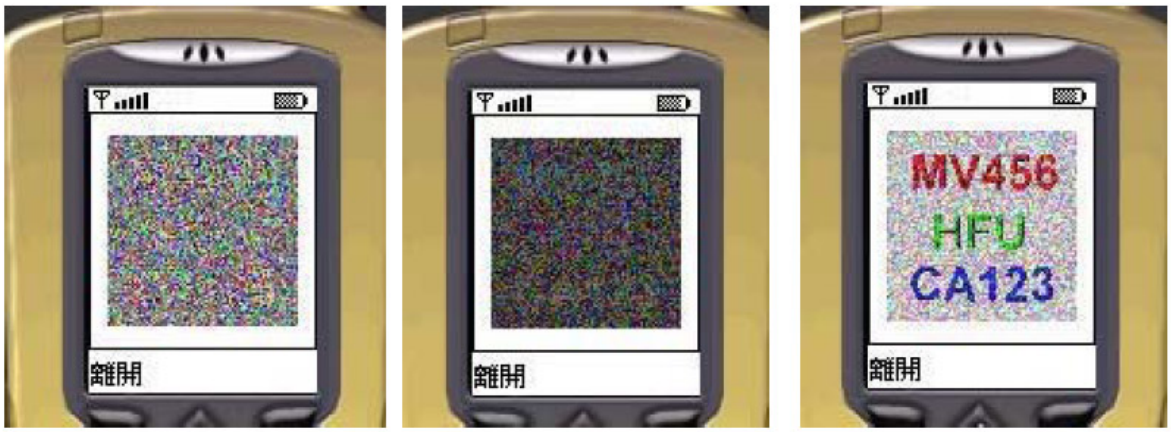


Abbildung 11.4: Mobile Ticket Validierung mit Zeitstempel [8]

diese Spaltung des Tickets in zwei Teile auch die Benutzerfreundlichkeit, weil das Ticket von Auge kontrolliert werden kann.

Ein weiteres erwähnenswertes Verfahren ist das schon angesprochene von ELCA [27]. Dieses Verfahren erlaubt die Authentifizierung der KonsumentInnen über ein Bild, das in das Ticket integriert wird (vgl. Abbildung 11.5). Es stützt sich dabei auf die Tatsache, dass in moderne Mobiltelefone eine Kamera eingebaut ist, mit welcher die KonsumentInnen ein Photo von sich machen müssen, um ein Ticket zu bestellen. Per MMS wird das Photo anschliessend an die Ticket-Applikation von ELCA geschickt, welche es durch ein patentiertes Mikrostrukturierungsverfahren verschlüsselt und zusammen mit einem 2D-Barcode, der die ticketrelevanten Informationen enthält, zurückschickt. Zwischendurch muss noch über 0900-Nummer, Premium Rate SMS oder Prepaidverfahren bezahlt werden. Solch ein SecuTix eignet sich also auch für personalisierte, nicht übertragbare Zutrittsberechtigungen in allen Bereichen des Ticketings. Nachteile lassen sich allerdings in der Handhabung und des eingeschränkten Kundenkreises erkennen, denn nicht jedes Mobiltelefon hat eine integrierte Kamera.


Abbildung 11.5: Die SecuTix-Lösung [27]

## 11.6 Fazit und Ausblick

Die Vielfältigkeit der Anwendungen und Anforderungen für mobiles Ticketing spiegelt sich in vielen unterschiedlichen Verfahren wider. Je nach Industriezweig und Ausrichtung der beteiligten Organisationen bringt das eine Verfahren mehr Vorteile als das andere. Gilt es, die KonsumentInnen weitestgehend zu befriedigen, sind einfache und automatisierte Abläufe wie z.B. eine automatische Abrechnung und RFID-Tickets besser geeignet, als lange dauernde und Konzentration fordernde Interaktionen wie z.B. bei programmbasierten Tickets. Sollen aber die Kosten für die Systemimplementation und damit auch für den Vorverkauf tief bleiben, wird eine Lösung die erste Wahl sein, die sich über günstige Netzwerke und Hardware abwickelt wie z.B. beim NoordNed-Verfahren. Wenn Sicherheit die Hauptrolle spielt, eignen sich z.B. das Zeitstempel- und das SecuTix-Verfahren, die Kopierschutz bzw. personalisierte Tickets ermöglichen. Steht aber die Kompatibilität und damit die Reichweite des Systems im Vordergrund, sind Lösungen gefragt, die über bereits vorhandene Schnittstellen jederzeit möglichst viele KonsumentInnen erreichen. Das Mobiltelefon erweist sich hier dank seiner Eigenschaft als persönliches, vertrauenswürdiges Gerät, das die KonsumentInnen häufig bei sich tragen, als geeignete Schnittstelle, die zeit- und ortunabhängige Bestellung, Versand, Abrechnung und Kontrolle erlaubt. Es wird sich zeigen, ob die KonsumentInnen die neuen Möglichkeiten, die ihnen die Mobilität bietet, wie angenommen nutzen. Auf der Seite der Anbieter dürfte die Motivation zur Integration mobiler Ticketing Systeme durch die möglichen Kostenersparnisse verstärkt werden. Insbesondere bei tiefen Ticketpreisen wie z.B. beim öffentlichen Verkehr ist das Kostenargument entscheidend. Schlussendlich werden sich diejenigen Lösungen durchsetzen, die sich am besten den jeweiligen Verhältnissen anpassen; Verfahren also, die den individuellen Anforderungen eines Szenarios gerecht werden, indem die Prozesse und der Technologiemix an die Bedürfnisse des Szenarios und insbesondere der KonsumentInnen angepasst werden.

# Literaturverzeichnis


[1] Mobile electronic Transaction: “MeT White Paper on Mobile Ticketing“, [online Dokument], Januar 2003: http://www.mobiletransaction.org/ pdf/R200/white_papers/MeT_White_paper_on _mobile_ticketing_v1.pdf

[2] T. Suman Kumar Reddy, H. Mohanty, R.K. Ghosh, S. Madria: “Two Distributed Algorithms for E-ticket Validation protocols for Mobile Clients“, 2003, IEEE

[3] K. Kuramitsu, T. Murakami, H. Matsuda, K. Sakamura: “TTP: Secure ACID Transfer Protocol for Electronic Ticket between Personal Tamper-proof Devices“, 2000, IEEE

[4] G. Me: “Security overview for m-payed virtual ticketing“, 2003, IEEE

[5] F. Bao, L. Anantharaman, R. Deng: “Design of Portable Mobile Devices Based E-Payment System and E-Ticketing System with Digital Signature“, 2001, IEEE

[6] F. Pedone: “Optimistic Validation of Electronic Tickets“, 2001, IEEE

[7] F. Bao: “A Scheme of Digital Tickets for Personal Trusted Device“, 2004, IEEE

[8] C. Chen, T. Lu: “A Mobile Ticket Validation by VSS Tech with Time-Stamp“, 2004, IEEE

[9] J. Franke: “Das Handy als Fahrkarte - Sicherheit für Mobile-Ticketing-Lösungen“, [online Dokument], 2003: http://www.competence-site.de/sbs.nsf/ 24ACFFCFF0072A15C1256D730021CBDF/ $File/f_fa_mobileticketing_010703.pdf

[10] T. Lerner, V. Frank: “Mobile Ticketing - Ticketverkauf für Busse und Bahnen über Mobiltelefone“, [online Dokument], 2003: http://www.businessvillage.de/Magazin/ mag_detail/mag-105_Mobile_Ticketing_-_Ticketverkauf_fuer_Busse_und_Bahnen 0Dube_Mobiltelefone.html

[11] Teltix: “Vier Wochen Mobile Ticketing in Osnabrück: Jeder fünfzigste Fahrschein kommt per SMS“, [online Dokument], 2003: http://www.teltix.de/pdf/presse_info_erste_zahlen.pdf

[12] BusinessVillage: “Mobile Ticketing: mehr Service - weniger Kosten“, [online Dokument], 2003: http://www.ecin.de/mobilebusinesscenter/mobileticketing/

[13] Sybase Magazine: “The Business Model Now Says ’Mobile Ticketing’“, [online Dokument], 2005: http://www.sybase.com/detail“id=1028621

[14] J. Verstaen: "Mobile Ticketing wird erwachsen", [online Dokument], 2004: http://www.competence-site.de/mbusiness.nsf/0/f8d0849235c1937fc1256ede00380d6e"OpenDocument

[15] Funkschau: "Das Handy als Ticket", [online Dokument], 2004: http://www.funkschau.de/heftarchiv/pdf/2004/04/fs_0404_s30.pdf

[16] A. Rossband: "Mobile Payment und Ticketing", [online Dokument], 2003: http://www.fh-coburg.de/fileadmin/fbeinf/staff/wieland/mobisys04/Postersession_Rossband.pdf

[17] LogicaCMG: "mobile ticketing - Ticketing solutions in the palm of your hand", [online Dokument], 2003: http://www.logicacmg.com/pdf/wes/mobile_ticketing.pdf

[18] mTicket: "mTicket Datasheet", [online Dokument], 2005: http://www.mticket.net/docs/mTicket.Datasheet.pdf

[19] RegiSoft: "RegiSoft's Mobile Ticketing", [online Dokument], 2002: http://www.regisoft.com/news/brochures/RegiSoft 20Mobile 20Tickting 20Feb 2002 20V1.pdf

[20] European Communities: "Innovative m-ticketing solution to be tested in Frankfurt", [online Dokument], 2004: http://europa.eu.int/idabc/en/document/3450/353

[21] ITS International: "World's first NFC enabled mobile product for contactless payment and ticketing", [online Dokument], 2005: http://www.itsinternational.com/News/article.cfm"recordID=7026

[22] Nokia: "Nokia kündigt weltweit erstes mobiles NFC-Produkt für kontaktlose Zahlungen und Ticketing an", [online Dokument], 2005: http://www.nokia.de/de/pressemitteilungen/nokia.de/2005/02/150038-framedPopup.html

[23] Gavitec AG: "Vodafone Load-a-Ticket - the mobile event ticket", [online Dokument], 2005: http://www.mobiledigit.de/fileadmin/template/main/downloads/CaseStudy2_Vodafone_LoadATicket.pdf

[24] SKIDATA AG: "LogicaCMG und SKIDATA: Mobiltelefon als Parkticket", [online Dokument], 2004: http://www.skidata.ch/company/news/LogicaCMG_SD_mobileticket.asp

[25] Accela Communications, Inc.: "Hamburg city pushes mobile phone ticketing", [online Dokument], 2004: http://wireless.itworld.com/4244/041126hamb/pfindex.html

[26] MATRIX Solutions GmbH: "kartenhaus.de startet mit PicTicket in die mobile Zukunft", [online Dokument], 2005: http://www.matrixsolutions.de/pressinformation.php"id=31

[27] SecuTix: "Mobile SecuTix - das MMS Ticket auf Ihrem Handy", [online Dokument], 2005: http://www.secutix.ch/resources/product_mobile_de.pdf

[28] UP-GREAT AG: “Smart e-Ticket - Das fälschungssichere e-Ticket“, [online Dokument], 2005: http://www.up-great.ch/wDeutsch/doc/produktsheet_ticket@home_v2_deu.pdf

[29] Siemens VDO Automotive AG: “3000 testen das modernste Ticket der Welt“, [online Dokument], 2005: http://www.siemens-tts.ch/haupt.asp“nv=2094&spr=1&ct=457

[30] VisionOne: “Erfolg mit elektronischem Ticketvertrieb“, 2005: http://www.visionone.ch/site/news02.asp“s=260

[31] Leuze electronic GmbH + Co KG: “Vergleich Data Matrix - Barcode“, [online Dokument], 2005: http://www.leuze.de/deutsch/datamatrix/www-vergleich-bc-dm.pdf

[32] Leuze electronic GmbH + Co KG: “Der Codeknacker“, [online Dokument], 2005: http://www.leuze.de/deutsch/datamatrix/www-allgemein-codeknacker.pdf

[33] K. Mattila: “Mobile Ticketing: Case Study Helsinki“, In: DVWG:“Echtzeitinformation & Mobiles Ticketing durch portable Verkehrstelematik: Erfahrungen und Innovationspotentiale“, 2003, S. 108-121

[34] G. R. Meijer: “Mobile Tickets “ Bahnfahren auf die neue Art“, In: DVWG: “Echtzeitinformation & Mobiles Ticketing durch portable Verkehrstelematik: Erfahrungen und Innovationspotentiale“, 2003, S. 122-128

[35] T. Bach: “Mobile Ticketing im ÖPNV - Fahrscheinverkauf über das Mobiltelefon in Osnabrück“, In: DVWG:“Echtzeitinformation & Mobiles Ticketing durch portable Verkehrstelematik: Erfahrungen und Innovationspotentiale“, 2003, S. 129-137

[36] H. P. Schär: “Automatische Anwesenheitserfassung im ÖPNV: Raumerfassung - Wohin geht die Reise““ In: DVWG:“Echtzeitinformation & Mobiles Ticketing durch portable Verkehrstelematik: Erfahrungen und Innovationspotentiale“, 2003, S. 138-151

[37] “LogicaCMG - Solutions That Matter“, [online Dokument], 2005: http://www.logicacmg.com/

[38] “Matrix Solutions“, [online Dokument], 2005: http://www.matrixsolutions.de/

[39] “Telecoms and Software Consulting“ [online Dokument], 2005: http://www.ovum.com/

[40] “NoordNed Personenvervoer B.V.“ [online Dokument], 2005: http://www.noordned-ov.nl/

[41] “Informa Telecoms and Media: research house providing the best business intelligence for the global wireless/fixed telecoms and media markets“ [online Dokument], 2005: http://www.informatm.com/

[42] “Vrije Universiteit Amsterdam“ [online Dokument], 2005: http://www.vu.nl

[43] “Kudelski Group - World leader in digital security“ [online Dokument], 2005: http://www.nagra.com/

[44] “SKIDATA Zutrittskontrollsysteme - Zutrittslösungen für Wintersportgebiete, Parkhausanlagen, Stadien, Messen und Freizeitanlagen“ http://www.skidata.ch/

[45] “Siemens Business Services GmbH & Co. OHG“ [online Dokument], 2005: http://www.sbs.de/

[46] “Siemens VDO Automotive AG“ [online Dokument], 2005: http://www.siemensvdo.com/

[47] “Nokia Schweiz“ [online Dokument], 2005: http://www.nokia.ch

[48] “Vodafone.com“ [online Dokument], 2005: http://www.vodafone.com/

[49] “Sony Schweiz“ [online Dokument], 2005: http://www.sony.ch/

[50] “Philips Country Selector page“ [online Dokument], 2005: http://www.philips.ch/

[51] “mTicket The original mobile ticket application“ [online Dokument], 2005: http://www.mticket.net/

[52] “HOME“ [online Dokument], 2005: http://www.gavitec.com/

[53] “Eventim.de - Der-Rund-ums-Ticket-Service - Konzertkarten einfach online bestellen¡‘ [online Dokument], 2005: http://www.eventim.de/

[54] “RegiSoft enables businesses to reach their consumers at the time and the place where the consumer is most receptive“ [online Dokument], 2005: http://www.regisoft.com/

[55] “TELTIX - Ihr Handyticket“ [online Dokument], 2005: http://www.teltix.de/

[56] “VisionOne - Worldwide“ [online Dokument], 2005: http://www.v-1.ch/

[57] “ELCA: Technology-Consulting-Innovation: job opportunities: careers: Informatiker: software engineer: Programmierer: jobs: software design: business strategy: crm: web design: data warehouse: edm“ [online Dokument], 2005: http://www.elca.ch/

[58] “UP-GREAT AG - the solution company“ [online Dokument], 2005: http://www.up-great.ch/

[59] “Swisscom - Startseite“ [online Dokument], 2005: http://www.swisscom.ch/

[60] “TicTec“ [online Dokument], 2005: http://www.tictec.ch/

[61] “ETH Zürich - Eidgenössische Technische Hochschule Zürich“ [online Dokument], 2005: http://www.eth.ch/

[62] “EPFL Entree principale“ [online Dokument], 2005: http://www.epfl.ch/

[63] “Good News - Homepage“ http://www.goodnews.ch/

[64] “DEAG Deutsche Entertainment AG - The World Of Entertainment“ [online Dokument], 2005: http://www.deag.de/

[65] “Ringier.CH“ [online Dokument], 2005: http://www.ringier.ch/

[66] “Freddy Burger Management“ [online Dokument], 2005: http://www.fbmgroup.ch/

[67] “Bienvenue sur Live Music Production“ [online Dokument], 2005: http://www.lmprod.ch/

[68] “SCHWEIZERISCHER FUSSBALLVERBAND“ [online Dokument], 2005: http://www.football.ch/

[69] “CIRCUS KNIE“ [online Dokument], 2005: http://www.knie.ch/

[70] “Opus One“ [online Dokument], 2005: http://www.opus1.ch/

[71] “Paléo Festival Nyon“ [online Dokument], 2005: http://www.paleo.ch/

[72] “Montreux Jazz Festival“ [online Dokument], 2005: http://www.montreuxjazz.com/

[73] “All Blues Konzert GmbH“ [online Dokument], 2005: http://www.allblues.ch/

[74] “Free & Virgin Group - Concerts, Artist Management, Events“ [online Dokument], 2005: http://www.freeandvirgin.com/

[75] “SBB“ [online Dokument], 2005: http://www.sbb.ch/

[76] “Startseite“ [online Dokument], 2005: http://www.sunrise.ch/

[77] “Willkommen bei Orange“ [online Dokument], 2005: http://www.orange.ch/

[78] “France Telecom - Bienvenue sur le site du Groupe France Télécom“ [online Dokument], 2005: http://www.francetelecom.com/

[79] “Official Ticketmaster site. Tickets for Concerts, Sports, Arts, Theater, Family, Events, more“ [online Dokument], 2005: http://www.ticketmaster.com/

[80] “starticket - print at home“ [online Dokument], 2005: http://www.starticket.ch/

[81] “Ticketcorner“ [online Dokument], 2005: http://www.ticketcorner.ch/

[82] “iAnywhere.com - Home“ [online Dokument], 2005: http://www.ianywhere.com/

[83] “www.kartenhaus.de“ [online Dokument], 2005: http://www.kartenhaus.de/

[84] “Fraunhofer-Gesellschaft“ [online Dokument], 2005: http://www.fraunhofer.de

# Chapter 12

# Technologies beyond 3G

*Dane Marjanovic, Philipp Buchmann*

*First generation networks provided us with simple voice telephony. Second Generation added the ability to send data like fax or email. The third generation provides data rates of up to 2 megabits per second and of course the conventional voice, fax and data service. This Speed allows us to have applications with high bandwidth like Videophone, TV feeds or home shopping. UMTS is such a 3G system. It offers a higher speech quality and data traffic. UMTS can be operated with the GSM networks. Because both systems use different frequency band, it should not interfere with each other. Many vendor are working further on WLAN-3G capabilities especially for roaming and networks handover. Having mentioned the continuous change and development in wireless mobile or fixed technology above, it is clear that existing 3G technologies and services are going to be further developed or maybe even replaced by new upcoming technologies. At present, however, there is no clear definition of what these technologies might be or how they are going to be implemented. What can be said is that new technologies keep emerging and that significant research and development is being done. By giving an overview of existing technologies, their requirements and usage possibilities we make an intro to new possible technologies and infrastructures that are emerging from existing trends and ongoing development. These new technologies are certainly not clearly defined yet, but we try nevertheless to point out and illustrate some technologies that are most likely to be implemented and considered as technologies of 4G. Carefully said, there might not at all be such a thing as 4G. But since this term is used even before (1G, 2G, 3G) we shall continue in this manner and provide an insight into 4G regarding new technologies to be implemented.*

## Contents

## 12.1 Technologies now - a short overview

At the beginning of this Paper, we will shortly go through all the common technologys we use today. We go from the architecture to the capacity and to the usage of the technology. After this overview of the technologies we use nowadays, we go farther to 4G. Devices, Protocols and how the whole thing will work will be disscused.

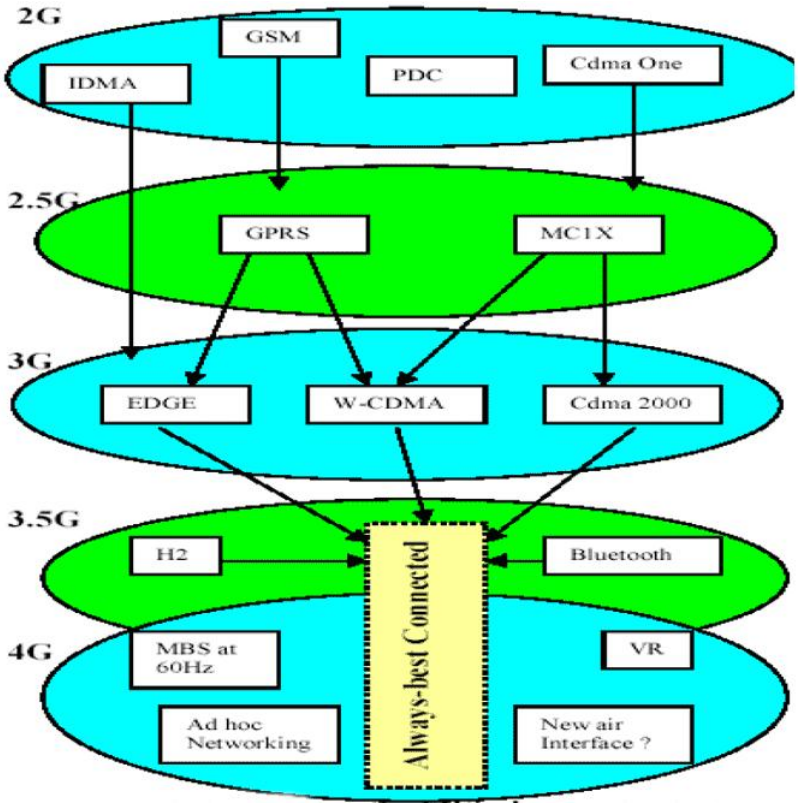


Figure 12.1: Upgrade path of wireless technologies

Most people uses GSM for voice calls as the main protocol. But what is with data transfer. Nowadays you need your mobile also for sending pictures and video (MMS), for going online with you notebook, while sitting in the train or checking emails while being in a café.

The need of higher bandwidth and the new applications needs a fast data transfer. At the moment you have the choice between GPRS (max. 57.6 kbits/s), HSCSD (max. 43.2 kbit/s) , CSD (9.6 kbit/s) and as the newest technology UMTS (384 kbit/s) connection. Today, most people uses GPRS, because of the high speed and the high availability. Usually, GPRS data is billed per kilobytes while circuit-switched data connections are billed per second. The latter is to reflect the fact that even during times, when no data is being transferred, the bandwidth is unavailable to other potential users. For an application such as downloading, HSCSD may be the better solution, since circuit-switched data is usually given priority over packet-switched date on a mobile network.

### 12.1.1 UMTS

#### Architecture

UMTS means Universal Mobile Telecommunications System and is one of the third-generation (3G) technologies. One of the requirements for this Architecture is the roaming possibility to the GSM System. UMTS is a combination of the protocol that defines over-the-air transmission between mobile phones and towers (W-CDMA), GSM protocol that allows to route calls from an to mobile subscribers (Mobile Application Part) and the GSM codecs that defines how audi has to be digitized. (AMR, EFR).

UMTS, from a existing GSM operater view, is a simple but costly migration from GSM. Most of their infrastructures remain the same, but they have to get licenses which are recently auctioned in Switzerland. Over 120 licenses have been awarded to mobile phone operators worldwide. As an example, in Germany alone, the license holders paid a total of 50.8 billion euros. Operators expect the ROI in 2005. UMTS is at the air interface level incompatible with GSM. The mobile phones sold nowadays are dual-mode phones (UMTS/GSM). A UMTS network includes 3 interacting domains:

- Core Network (CN)
- UMTS Terrestrial Radio Access Network (UTRAN)
- User Equipment (UE)

The main function of the UMTS network is to provide routing, switching and transit for the user traffic. The Core network is based on a GSM network with GPRS capabilities. The Network hast to know approximate the location of the user in order to be able to page user equipment. Here's a list of areas from smallest to largest.

- Sub cell
- Cell
- UTRAN Registration Area (PS Domain)
- Routing Area (PS Domain)
- Location Area
- MSC/VLR or SGSN
- Public Land Mobile Network (PLMN)
- UMTS systems

The Cs-Domain is the circuit switched domain, which provice services like speech calls. The PS-Domain is the packet switched domain, which provide services such as IP Based traffic.

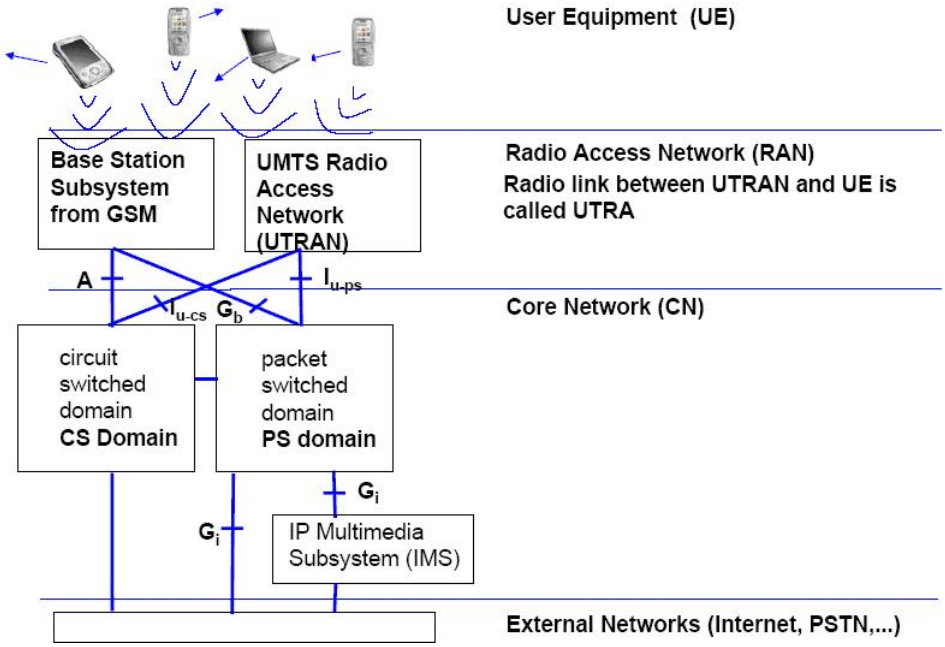


Figure 12.2: UMTS basic network architecture

### Capabilities

UMTS represents an evolution in the data speed and broadband capabilities to support greater numbers of customers (voice and data). UMTS is ideally suited for applications with the need of higher bandwidth like videoconference or livetv feed's. Except for the higher bandwidth, UMTS offers enhanced capabilities in terms of Security, QoS ec. UMTS offers also a significant higher capacity and broadband capabilities to support greater number of data and voice customer. This is especially important in urban centres. UMTS supports up to 1920 kbit/s data transfer rates, although users in the real networks can expect a transfer rate of 384 kbit/s at the moment. However, this is much greater than the good old GSM Circuit Switched data channel (14.4 kbit/s) or GPRS (56 kbit/s).

E-GPRS or EDGE is a further evolution of GPRS. It's based on a new coding schemes. EDGE is also often called as 2.75G-Systems. The Bandwidth of EDGE is around 180kbit/s In the future it's possible to upgrade todays UMTS with High Speed Downlink Packet Access (known as 3.5G). This would increase the downlink transfer to 14.4Mbit/s

### Usage

Applications and services, which takes advantage of the UMTS Capabilities are expected in the next few years. Web Browsing, multimedia, live tv feeds and video conferencing are just few examples. Swisscom for example, already serve liveTV (SF DRS: 1 and 2, SF INFO, MTV, Eurosport, TSR 1 and 2, TF 1, TSI 1 and 2) are 24/7 live online and you can watch it for 1.50 CHF per hour or 4 CHF per day. Videoconferencing is also one of the killer application at the moment.

### 12.1.2 GPRS

#### Architecture

General Packet Radio Service (GPRS) is the nowadays the most used data service. GPRS is also often called 2.5G, what means a technologie between the second and the third generations of mobile phones. GPRS has a moderate bandwith of 30-70kbit/s by using unused channels in the GSM network. GPRS is, other than GSM, packet-switched which means that multiple users share the same channel. That means, that the total available bandwith can bi dedicated to those users who sending or receiving at any given moment. Receiving eMails, browsing the web and instant messaging are examples of uses which needs intermittent data transfers and are best suited for the use of GPRS.

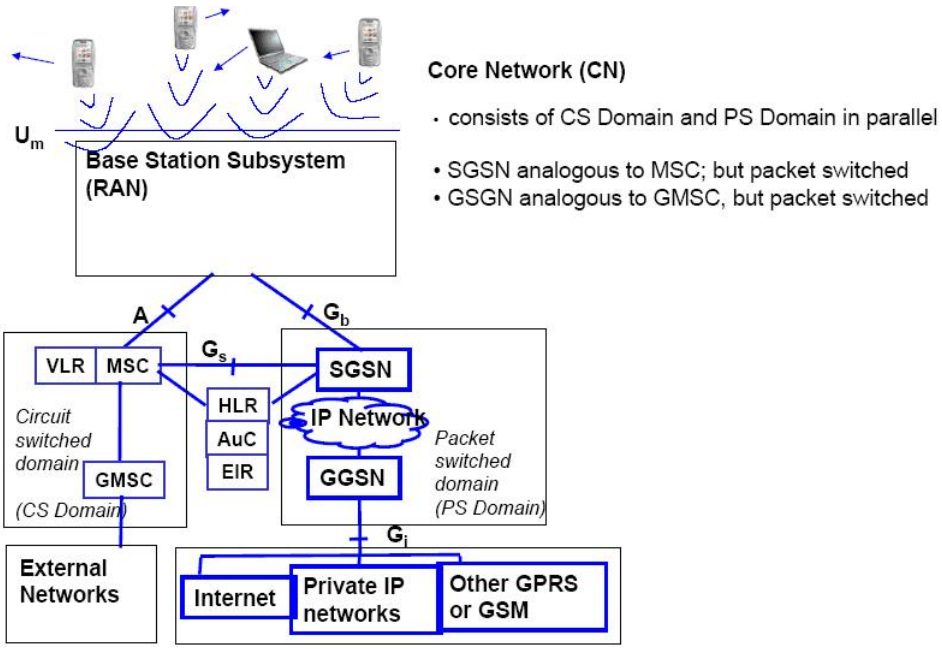


Figure 12.3: Simplified GSM architecture

#### Capabilities

More efficient usage of resources statistical multiplexing All application share the same resource. This is especially efficient for applications with variable rates. GPRS have a higher bandwidth than GSM (Max 171.2 kb/s).The Packet switched GPRS system allows the operator a new pricing model: charging per date volume. Swisscom for example charges 0.10 CHF per 10kb (10 CHF / MB). This allows us to stay always online without paying online time. You pay only for the traffic you make.

A other advantage is the shared bandwidth, which allows the operator to use the full capacity of his network. The handshake is also very fast in comparison to GSM.

#### Usage

Every application which needs online access can make usage of the GPRS protocol. Checking emails, surfing the WAP portals or update the actual traffic information with your Mobile Navigation System (ex. TomTom Mobile) is just a little selection of many applications. Most applications which needs to stay online but doesn't make traffic without user intervention are ideally suited for the packet-switched GPRS protocol. Because of the faster handshake and the higher bandwidth are MMS also sent through GPRS.

### 12.1.3 WLAN

#### Architecture

There are 2 types of Wi-Fi network types:

**Infrastructure** This Wi-Fi Network type requires at least one access point being connected to a LAN. This setup will allow WLAN devices to make use of the resource on the LAN such as accessing files, printers and of course the internet.

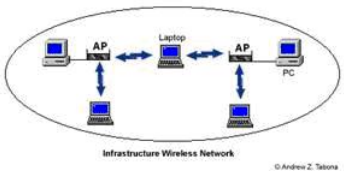


Figure 12.4: Infrastructure WLAN network

**AdHoc** On this Wi-Fi network type, the clients connect peer to peer to each other. No access point is needed.


Figure 12.5: adhoc WLAN network

**WEP: Wired Equivalent Privacy**
It is suitable for most small networks and is intended to stop the interception of radio frequency signals by unauthorized users. Because of the manually entering of the Key in every client, this is a very time consuming task. Wep encryption is based on RC3 by RSA Data System.

**SSIS: Service Set Identifier**
SSID allows a simple password to split the WLAN Network in different networks. Every

Access Point has a own SSID and every client has the appropriate SSID aswell. If the SSID Match (Client <-> AP), access will be granted.

**MAC: Media Access Control address filtering**
A user defined list in the Access Point can be inputted to grand them access. Evertime a Client makes a request, the MAC address will be compared to the MAC Adress list in the AP and permission will be granted or denied.

**VPN: Virtual Private Network Link**
This is perhaps the most secure form of a Wi-Fi setup. The Client make a secure tunnel (over the WLAN) to the VPN Server. This Tunnel is for example secured by IPSEC which makes decrypting the traffic a very hard, if not impossible task.

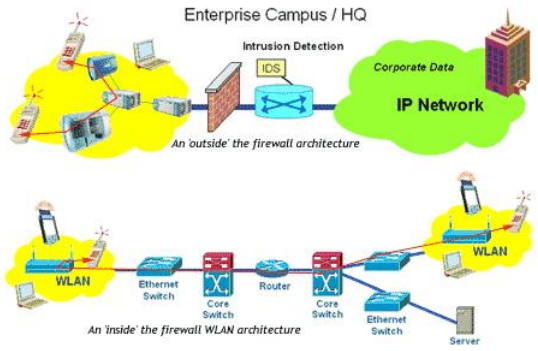


Figure 12.6: VPN Architecture inside and outside of the firewal

## Capabilities

One big advantage of Wi-Fi is, that unlike packet radio systems, Wi-Fi uses unlicensed radio spectrum . So it doesn't need regulatory approval. High data rates (up to 54mbs), low cost and a easy installation makes WLAN to an impressive technologie. Competition amongst vendors lowered the prices and makes this technology very advantageous. Some argue that Wi-Fi has the possibility to replacing cellular phone networks such as GSM. But Wi-Fi has some handicaps in missing roaming and authentication, the narrowness of th available spectrum and the limited range of Wi-Fi. Regardless of the problems, companies like Zyxel, SocketIP are already offering telephony platforms that use Wi-Fi transport. Many operators are selling mobile products that combine cellular wireless and Wi-Fi radio system in a more or less transparent way to take advantage of the benefits of both technologies.

The Wi-Fi networks can differentiate between Free and commercial Wi-Fi networks. The commercial services are available in places such as Internet Cafes, airports, hotels or on any other public places. While the Free Wi-fi attempt is to share the own Wi-Fi connections with others so everybody can use and benefit for the fast and free Internet Access. Many community groups have built their Wi-Fi networks entirely based on volunteer efforts and donations. Many Universities (also UNI ZH) provide free Wi-Fi internet access to their students on the whole campus.

**Usage**

The Usage of a WLAN is ernormous but als dangerous. The fact that people aren't using encryption and password protection is dangerous but it depends what somebody can do, if if he's on the network. If the only possibility is to browse the internet it's not so dramatic. But if the company's network drive it's also accessible it can be very disastrous. A network administrator has the duty to know the risks and know how they can be eliminated or at least restricted. It exists different approaches from a normal MAC Adress filter, WEP ciphering to a VPN gateway in IPSEC. WLAN is great for mobile computing. Taking Surfing the net while waiting for the flight or checking emails while sitting in the lobby are great potentials of this technology.

## 12.2 Migration considerations from 3G to 3.5G and 4G

### 12.2.1 Infrastructure changes

On average the cell size are smaller than 3G. However, the cell size will be made as large as possible via high power base stations, asymmetriy and adaptive antennas option.

### 12.2.2 Communication protocol changes

The 4G technology can be imagined as a integrated wireless system. It will allows us to have seamless roaming between the different technologies. It doesn't matter if the user operate in a cellular technology or on a wireless network. The handover will be smoothly and depends only about the charging preference. The protocol will be entirely packet switched. In 3G exists both: circuit and packet switched networks. The IP protocol will be based on IPv6.

### 12.2.3 New devices

The design of a user Terminal, that can operate in different wireless networks, is difficult. The limitation in the device size, cost, power consumption as well as the backward compaibilites to the old systems will make major problems.

## 12.3 From 3G to 4G

This Chapter covers some interesting topics regarding the migration from 3G to 4G. In times where new technologies emerge and are considered to be the future technologies in

use, one question always comes up: What technologies are going to prevail and which one of them is really going to come to use? This chapter focuses on the answer to this question. First we shall look a bit deeper into the discussion of what mobile wireless technology is becoming. We will mention some possible considerations about where mobile technology is going, meaning what 4G should be all about. Then we'll examine some possible trends in development of future mobile wireless technology. These trends however are mostly initiated by industry companies that have great interest an strong incentives in investing in such technology.

### 12.3.1 Where is mobile technology going?

The answer to that question is obviously not so straight forward. However there is great discussion on what 4G is really meant to be. Is it a conceptual framework to address the future needs for a high speed wireless network or is it considered to be a set of new transmission technologies, protocols, maybe devices to mach the grooving demand for higher bandwidth and speed in wireless networks?

From our point of view 4G can be imagined as an integrated wireless system that enables seamless roaming between technologies. Hereby a user can be operating in cellular technology network and get handed over to a satellite-based network and back to a fixed wireless network, depending upon the network coverage and preference of charging. This integrated wireless system should be based, among other, on deeper focusing on user experience. This is straightforward, because when we look at the so called 3G story we see that the technological considerations were centered mainly on making a new, more capable, air interface. There was significant over expectation however. Further the research around 4G should be cantered around architecture and system aspects that would support an effective, open, flexible integration of multiple technologies.

4G technology should provide support for interactive multimedia services such as teleconferencing, wireless internet, etc. On the more technical side is should provide wider bandwidths and higher bit rates. Further there has to be a global mobility for users and seamless and easy service portability at a relatively low cost. The low cost argument goes for reducing infrastructure and providing costs by network providers as well. After the above considerations here we shall provide a rather insightful graph that takes into consideration even more business related issues but strongly bound to the emerging new technologies.

The picture illustrates some further technologies and services that can emerge within 4G. Hereby the two main factors are, on one hand to increase the scope of use and application opportunities for users and on the other providing an increasing bandwidth and capacity of carriers. Increasing the scope of use could ergo lead to the real "Anything, Anytime, Anywhere" experience that is a main driver of 4G considerations. Increased bandwidth and carrier capacity leads us into 4G technologies infrastructures and services and their further development.

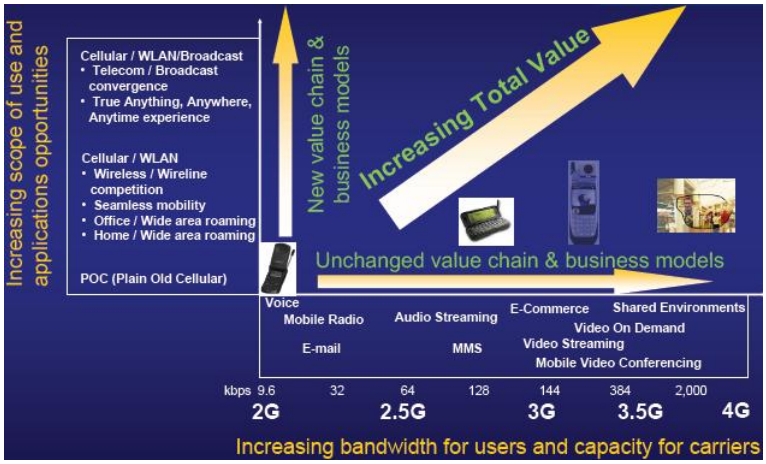


Figure 12.7: Picture1: Mohammad Abualreesh, presentation on 4G

### 12.3.2 Possible trends

So far we have looked at the possible definitions of 4G technologies and we put up a somewhat clear and straightforward definition of what 4G should be. Though there isn't only one definition and only one trend towards 4G. This is why in this subsection we try to cover some possible trends initiated by large telecommunication firms and consortia around the problematic of 4G technologies. Trends are not just future visions of technology but almost clearly defined ways to embrace, develop and deploy new emerging technologies.

One of the major trends toward 4G is the great heterogeneity of the deployed networks. Given both investors' requirements for capital expenditure intensity and the technological divergence of private and public networks, the issues of service portability and interoperability have become of primary importance.

Another trend, according to the discussion on "Topics in 4G Wireless Networks: Ad Hoc Nets, Adaptive Services and QoS" recently initiated, are novel concepts for promoting opportunistic and adaptive formation of wireless network infrastructures as a way of overcoming critical technical and business model limitations of today's cellular and WLAN networks. The proposal here was to create a wireless "network of networks" together with techniques for supporting heterogeneous radios, protocols for self organization and ad hoc routing, and services that adapt to dynamic and opportunistic exploitation of available or expected resources to meet application needs.

Always Best Connected (ABC), a fixed and mobile wireless access vision is another trend to be considered. It goes beyond the "always-on" slogan and seeks a firm foundation in already existing, expected, proposed, and yet to be conceived networks. The trend is perceived as a strategic goal to inform and define the significant advances happening and forecast in technologies, networks, user terminals, services, and service delivery, and in the future business models encompassing these, which will help crystallize and realize the 4G paradigm.

So we looked at a few possible trends towards 4G. At this point it should be said that

these and other trends we have examined in our research have some of the considerations in common. First there is the clear issue about having an open and heterogeneous and open infrastructure, then the possibility of seamless service and network switching and of course the increased availability and speed of new networks and infrastructures. In the following section we shall look at a few more significant industry initiatives in moving towards 4G.There we look a bit closer at the efforts of industry firms in reaching 4G:

**Industry initiatives regarding 4G**

The is first the WWRF (Wireless World Research Forum) consisting of Nokia, Alcatel, Siemens and Ericsson. Further on there is NTTDoCoMo, who has started conceptual designs for 4G. The objective of this forum is to formulate visions on strategic future research directions in the wireless field, among industry and academia, and to generate, identify, and promote research areas and technical trends for mobile and wireless system technologies. It is also intended to closely co- operate with the UMTS Forum, ETSI, 3GPP, IETF, ITU, and other relevant bodies regarding commercial and standardisation issues derived from the research work.

In keeping with its global leader-ship position, Samsung has taken a visionary stance on establishing 4G mobile communications initiatives. Samsung is interested in building relationships across industry and academia to further the efforts toward 4G standardization and development. Samsung is also interested in investigating cost-effective practical applications of 4G. As a beginning step toward nurturing a 4G standard, Samsung has been encouraging relationships among a wide range of interested commercial and academic partners.

Well, so far there has been a lot word on what 4G is meant to be and what it should or will incorporate and become. There have been also trends talked of, that have, among other, a great influence on the new mobile wireless technology. The next chapters are conceived in a more technical fashion. The discussion above should be regarded as an intro to the following chapters. It was intended to give the interested reader an insight and short overview of 4G technology and considerations regarding it.

## 12.4 4G infrastructures and technologies

Now that we have examined 3G Technologies and looked at some possible trends in mobile wireless technology it's time to look at technologies that are most likely to represent the core technologies of the new 4G mobile and wireless infrastructure. Though it is not clearly defined, which technologies are going to prevail or to be enhanced or even newly created, we try to provide a set of technologies that are, in our eyes, the technologies that will or should be migrated into and newly designed for the 4G infrastructure and services. According to this the technologies concerning network access and transmission discussed here will be the following:

- Wi-Max (Worldwide Interoperability for Microwave Access)

- OFDM; MC-CDMA, TDMA (Orthogonal Frequency Division Multiplexing, Multi Carrier- CDMA, )
- MIMO-OFDM (Multiple Input, Multiple Output- OFDM)

Obviously there is a great deal of FDMA (Frequency division multiple acces) technology. This is because we see this technology and its different versions become a very important core part of 4G infrastructure besides Wi-MAX, witch is taking more and more presence on the wireless communication stage and eventually it will succeed Wi-Fi as a wireless internet technology.

The presentation and discussion of the above mentioned technologies is to be considered with caution. Eventually some existing technologies might be changed or some new technologies will be introduced. It is not our intention to fantasize about new and cool technologies that may, or may not, emerge, someday, in a galaxy far away but rather to give a somewhat precise description of real existing technologies that have a high probability of being effectively used. Having said that, it's clear that sometimes we have to deal with technologies already deployed in 3G but that might be reused or modified for 4G usages. Never the less, we shall now look at each technology at the time.

### 12.4.1 Wi-MAX

Wi- MAX stands for: Worldwide Interoperability for Microwave Access It refers to any broadband wireless access network based on the new IEEE 802.16 standard. The smallest-scale network is a **personal area network** (PAN). A PAN allows devices to communicate with each other over short distances. Bluetooth is the best example of a PAN. The next step up is a **local area network** (LAN). A LAN allows devices to share information, but is limited to a fairly small central area, such as a company's headquarters, a coffee shop or your house. Many LANs use WiFi to connect the network wirelessly. WiMAX is the wireless solution for the next step up in scale, the **metropolitan area network** (MAN). A MAN allows areas the size of cities to be connected

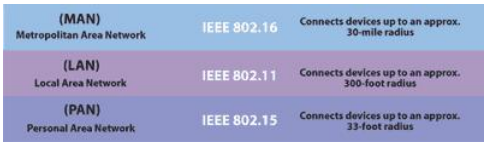


Figure 12.8: What Wi-MAX is suited for

#### Technical intro

In this part a technical overview of Wi-MAX is given. In clear terms this means we look at the technical issues such as the way it works (protocols, connection logic...) and give some information of the capabilities in terms of speed, bandwidth, delay, antenna range an so on.

Well, in practical terms, WiMAX would operate similar to WiFi but at higher speeds, over greater distances and for a greater number of users. Just as WiFi it sends data from one computer to another via radio signals. A computer (either a desktop or a laptop) equipped with WiMAX would receive data from the WiMAX transmitting station, probably using encrypted data keys to prevent unauthorized users from stealing access.

IEEE 802.16 (Wi-Max) specifications:

- Range: 30-mile (50-km) radius from base station
- Speed: 70 megabits per second
- Line: of-sight not needed between user and base station
- Frequency bands - 2 to 11 GHz and 10 to 66 GHz (licensed and unlicensed bands)
- Defines both the MAC and PHY layers and allows multiple PHY-layer specifications

A Wi-max system consists of two parts:

- A **Wi-max tower**, similar in concept to a cell-phone tower - A single WiMAX tower can provide coverage to a very large area – as big as 3,000 square miles ( 8,000 square km).
- ” A **Wi-mac receiver** - The receiver and antenna could be a small box or PCMCIA card, or they could be built into a laptop the way WiFi access is today.

Figure 12.9: Wi-MaX antenna

A WiMAX tower station can connect directly to the Internet using a high-bandwidth, wired connection (for example, a T3 line). It can also connect to another WiMAX tower using a line-of-sight, microwave link. This connection to a second tower (often referred to as a backhaul), along with the ability of a single tower to cover up to 3,000 square miles, is what allows Wi-max to provide coverage to remote rural areas.

Wi-max can actually provide two forms of access (see also picture 4):

- ” The **non-line-of-sight**, WiFi sort of access, where a small antenna on your computer connects to the tower. In this mode, Wi-max uses a **lower frequency range** – 2 GHz to 11 GHz (similar to WiFi). Lower-wavelength transmissions are not as easily disrupted by physical obstructions – they are better able to diffract, or bend, around obstacles.

- " The **line-of-sight** service, where a fixed dish antenna points straight at the Wi-max tower from a rooftop or pole. The line-of-sight connection is stronger and more stable, so it's able to send a lot of data with fewer errors. Line-of-sight transmissions use **higher frequencies**, with ranges reaching a possible 66 GHz. At higher frequencies, there is less interference and lots more bandwidth.


Figure 12.10: WiMax access possibilities and possible infrastructure

### Infrastructure requirements

This part covers the infrastructures required for Wi.max such as antennas, receivers, etc. The Problematic of the infrastructures is one of the main dispute factors over the question: will Wi-MAX replace Wi-Fi permanently? Wi-MAX uses fixed network infrastructures and requires relatively high-gain antennas. The infrastructure requirements for Wi-max are quite modest. So far we have seen, that Wi-max basically operates the same way as WiFi so no greater transmission protocol changes ought to be made. Concerning "'hardware"', we have seen that for Wi-max we require a somewhat cell-phone tower that can provide coverage to a large area ( 8000m2). Also new Wi-max receiving devices have to be introduced but these devices will be similar to today's routers, accesspoints and PCMCA Cards and eventually Wi-max is going to be integrated as a communication technology directly in mobiles or portables. Yet, some new problems arise and have to be considered. For instance, the fact that base station sare now capable of covering wider areas, brings up the problem of being able to serve more users from a single base station. There is the risk of not being able to deliver service to users that are at a greater distance from the base station because of greater user concentration near the basestation.

### Usage possibilities

Here we look at the possible usage of Wi-MAX considering its functionality and infrastructure requirements.

In general, Wi-MAX is ideally suited to replace the costly installations and service of short-distance dedicated enterprise T-1 lines. Possible usage also goes into wirelessly connecting different buildings to a campus WAN

But, when we look at it a bit deeper we see that there are two possible ways Wi-max can be implemented – as a zone for wireless connections that single users go to when they want to connect to the Internet on a laptop (the non-line-of-sight "super WiFi" implementation), or as a line-of-sight hub used to connect hundreds of customers to a steady, always-on, high-speed wireless Internet connection.

Some companies might set up Wi-max transmitters and then make people pay for access. Again, this is similar to strategies used for WiFi, but a much wider area would be covered. Instead of hopping from one hot spot to another, Wi-max-enabled users could have Internet access anywhere within 30 miles of the Wi-max base station. These companies might offer unlimited access for a monthly fee or a "pay as you go" plan that charges on a per-minute or per-hour basis.

A short scenario might illustrate the usage of Wi-max a bit further (taken from `http://computer.howstuffworks.com/wimax.htm`):

An Internet service provider sets up a WiMAX base station 10 miles from your home. You would buy a WiMAX-enabled computer (some of them should be on store shelves in 2005) or upgrade your old computer to add WiMAX capability. You would receive a special encryption code that would give you access to the base station. The base station would beam data from the Internet to your computer (at speeds potentially higher than today's cable modems), for which you would pay the provider a monthly fee. The cost for this service could be much lower than current high-speed Internet-subscription fees because the provider never had to run cables. If you have a home network, things wouldn't change much. The WiMAX base station would send data to a WiMAX-enabled router, which would then send the data to the different computers on your network. You could even combine WiFi with WiMAX by having the router send the data to the computers via WiFi.

### 12.4.2 OFDM; MC-CDMA, TDMA

The reason, why we emention OFDM might not be so straightforward, since OFDM is rather a transmission protocol than an actual framework or infrastructure. Well, taking the fact thaat the idea of frequency multiplexing is nearly twenty years old, OFDM (FDM in general) is a wellconcieved and visionary protocoll. Due to implementation impossibilities in recent years, OFDM was not so broadly deployed but with improvements OFDM might become a dominant transmission protocol for future 4G networks.

In general, frequency division multiplexing (FDM) is a technology that transmits multiple signals simultaneously over a single transmission path, such as a cable or wireless system. Each signal travels within its own unique frequency range (carrier), which is modulated by the data (text, voice, video, etc.).

**Technical Intro**

Orthogonal Frequency Division Multiplexing is an FDM modulation technique for transmitting large amounts of digital data over a radio wave. OFDM works by splitting the radio signal into multiple smaller sub-signals that are then transmitted simultaneously at different frequencies to the receiver.

However one underlying technology is MCM (Multi carrier Modulation). Two different types of MCM are likely to be candidates for 4G: multicarrier code division multiple access (MC-CDMA) and orthogonal frequency division multiplexing (OFDM) witch we mentioned above, using time division multiple access (TDMA). Note: MC-CDMA is actually OFDM with a CDMA overlay. In OFDM with TDMA, the users are allocated time intervals to transmit and receive data. As with 3G systems, 4G systems have to deal with issues of multiple access interference and timing.

As you can see, there is plenty of technologies around FDM! Here is a small picture that illustrates the functionality of OFDM (FDM in general):

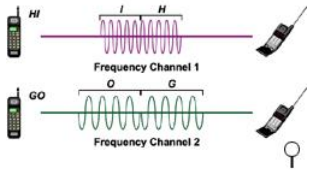


Figure 12.11: OFDM functionality

### 12.4.3 MIMO-FDMA

Multiple Input, Multiple Output Orthogonal Frequency Division Multiplexing is a technology developed by Iospan Wireless that uses multiple antennas to transmit and receive radio signals. MIMO-OFDM will allow service providers to deploy a Broadband Wireless Access system that has Non-Line-of-Sight functionality. Specifically, MIMO-OFDM takes advantage of the multipath properties of environments using base station antennas that do not have LOS The MIMO system uses multiple antennas to simultaneously transmit data, in small pieces to the receiver, which can process the data flows and put them back together. This process, called spatial multiplexing, proportionally boosts the data-transmission speed by a factor equal to the number of transmitting antennas. In addition, since all data is transmitted both in the same frequency band and with separate spatial signatures, this technique utilizes spectrum very efficiently.

**Infrastructure requirements**

Now, based on the technical and functional characteristics of the above mentioned technologies, we shall briefly look at the infrastructural requirements that must be met in order to make use of these technologies.

Since OFDM and its versions are mostly protocols that are used by other, in our case, wireless technologies, the infrastructure requirements fall together with the requirements for these wireless technologies.

However, there has to be the proper interfaces and implementations for OFDM and the like in order to fully and effectively make use of them.

**Usage possibilities**

Again we look at some possible usage of the above mentioned technologies considering their functionality and infrastructure requirements. One of the interesting things about OFDM is its relatively broad use and thus flexibility. Possible usues:

- DAB - OFDM forms the basis for the Digital Audio Broadcasting (DAB) standard in the European market.
- ADSL - OFDM forms the basis for the global ADSL (asymmetric digital subscriber line) standard.
- Wireless Local Area Networks - development is ongoing for wireless point-to-point and point-to-multipoint configurations using OFDM technology
- In a supplement to the IEEE 802.11 standard, the IEEE 802.11 working group published IEEE 802.11a, which outlines the use of OFDM in the 5.8-GHz band.

## 12.5 Outlook

Now that we have examined the technologies and looked at some possible trends in wireless mobile communication it's time to look a bit deeper into the future We do this by first making a summary of what is needed to build 4G networks of future :

- Lower price points only slightly higher than alternatives
- More coordination among spectrum regulators around the world
- More academic research
- Standardization of wireless networks in terms of modulation techniques, switching schemes and roaming is an absolute necessity for 4G

- A voice independent business justification thinking
- Integration across different network topologies
- Non disruptive implementation: 4G must allow us to move from 3G to 4G

The next step in our outlook is to examine some possible applications of 4G. Let's do so by listing just a few:

- Virtual Presence: 4G system gives mobile users a "'virtual presence"' –for example, always-on connections to keep people on event.
- Virtual navigation: A remote database contains the graphical representation of streets, buildings, and physical characteristics of a large metropolis. Blocks of this database are transmitted in rapid sequence to a vehicle.
- Tele-medicine 4G will support remote health monitoring of patients.
- Tele-geoprocessing Queries dependent on location information of several users, in addition to temporal aspects have many applications.
- Crisis-management applications
- Education

If there is to be a major, a "'killer"' application for 4G remains open. But what can be said from our point of view is that this application will definitely be in the mobile world providing wireless internet access and mobile services. What kind of services can be provided is still to be discussed.

## 12.6 Conclusion

Let us make a final conclusion on the topic of Technologies beyond 3G.

Concerning 4G, following is to be said:

- 4G can be imagined of as an integrated wireless system that enables seamless roaming between technologies.
- A promising 4G can support interactive interactive multimedia services with wider bandwidths, and multimedia services with wider bandwidths, and higher bit rates
- 4G still to come

# Bibliography


[1] The official 4G site, www.4g.co.uk

[2] www.networkworld.com/columnists/2004/0112wizards.html

[3] www.item.ntnu.no/fag/tm8100/Pensumstoff2004/UMTS2ANNE.pdf

[4] www.commdesign.com/story/OEG20010626S0065

[5] z2.inf.tu-dresden.de/andy/Zusatzfolien/InRichtung4G.pdf

[6] www.mobileguru.co.uk/MobileTechnologyglobe.html

[7] www.zzz.com.ru/zzzoriginalsite/art61.html

[8] www.tkn.tu-berlin.de/curricula/ws0405/vl-umts/UMTSArchitectureIII0405.pdf

[9] www.comlab.hut.fi/opetus/333/20042005slides/4G.pdf

[10] www.umtsworld.com

[11] www.windowsnetworking.com/articlestutorials/Introduction-Wireless-Networking-Part2.html