Mobile Systems III

Burkhard STILLER
Frank EYERMANN
Arnd HEURSCH
Peter RACZ
(Hrsg.)

Institut für Informationstechnische Systeme, IIS

Bericht Nr. 2004-02
Juni 2004

Universität der Bundeswehr München

Fakultät für

INFORMATIK

Werner-Heisenberg-Weg 39 • D-85577 Neubiberg

B. Stiller, F. Eyermann, A. Heursch, P. Racz (edts.): Mobile Systems III
Technical Report No. 2004-02, June 2004
Information Systems Laboratory IIS, Departement of Computer Science
University of the Federal Armed Forces Munich
Werner-Heisenberg-Weg 39, D-85577 Neubiberg, Germany
URL:http://www.informatik.unibw-muenchen.de/reports/#2004

Introduction

The Information System Laboratory (Institut für Informationstechnische Systeme, IIS) of the Department of Computer Science, University of the Federal Armed Forces Munich, Germany offered its students a new lap of the sequence of seminars on 'Mobile Systems', this time during the winter term 2004 (WT04).

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 identication 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 third edition of this seminar - entitled 'Mobile Systems III' - discusses in the first section 'War Driving'. While the search for open, unsecured, and publicly available Wireless Local Area Networks (WLAN) occurred to be a 'sports event', today a well established group of people keep track of those activities. This talk outlines the range of actions, goals of these activities, and the environment in which these people operate.

The second section addresses the area of secure networking across public networks. In particular it lists and discusses known Virtual Private Networks (VPN) as well as IPSec. Their comparison and use in wired and wireless networking domains conclude this section.

The third section is driven by demands of real-time applications, which require a real-time networking infrastructure. Therefore, the key aspects of real-time support of wired and wireless networking devices are discussed and major mechanisms in support of those demands are compared.

Due to the existence of wireless networks the support of mobility has changed tremendously the usage of services of today's Internet users. Therefore, the fourth section focuses in more detail on this area, in particular on micro-mobility support in IP-based networks

as well as Mobile IP extensions in terms of hand-off optimization and paging. A quite extensive comparison of existing protocols concludes this section.

The fifth section deals with QoS issues in the WLAN environment and outlines respective standards available in LANs. This includes a discussion of IEEE 802.1p and IEEE 802.1Q as well as a definition of QoS and one possible metering framework for QoS, the Real-Time Flow Measurement (RTFM) work.

Driven by possible solutions and key security concerns in mobile networks, section six develops a view on advanced concepts for security in those ones. The standard IEEE 802.11i addresses wireless security mechanisms, which are analyzed and advantages as well as drawbacks are discussed. Protocol-independent security measures are included in this section as well.

Due to the current security debate, section seven deals with intrusion detection systems, in particular addressing mobile environments. Key functionality is described and methods applied are discussed, while a view on wireless systems is being made.

Finally, section eight presents an overview on software environments for mobile devices. The key role plays the Operating System (OS), and major requirements in a mobile environment are discussed. Existing OS are presented and evaluated. Finally, J2ME is outlined as a programming environment for mobile systems.

Seminar Operation

As usual and well established now, all interested students worked on an initially offered set of papers and book chapters, relating to the topic titles as presented in the Table of Content below. They prepared a written essay as a clearly focussed 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 student 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 IIS support for preparing talks, reports, and their preparation by students had been granted Frank Eyermann, Peter Racz, Arnd Heursch, and Burkhard Stiller. A larger number of pre-presentation discussions have provided valuable insights in the emerging and moving field of Mobile Systems, both for students and supervisors. Many thanks to all people contributing to the success of this event, which has happened again in a small group of highly motivated and technically qualified students and people.

Neubiberg, May 2004

Inhaltsverzeichnis

1 War Driving - An Approach to Locate Open WLANs **7**
Dirk Schindler

2 Neue VPN-Lösungen **31**
Atnarong Kongnin

3 Realtime Networking in Wireless and Wired Networks **53**
Ronny Nantke

4 Micro-Mobility in IP-based Networks **77**
Philipp Appelhoff

5 Quality of Service in Wireless Local Area Networks **105**
Andreas Fischer

6 Moderne Konzepte für Sicherheit in Mobilen Netzen **121**
Jonathan Albe

7 Intrusion Detection in mobilen Netzwerken **147**
Lars Langer

8 Software Environments for Mobile Devices **165**
Michael Böhm

Kapitel 1

War Driving - An Approach to Locate Open WLANs

Dirk Schindler

Im Zeitalter von Information und Kommunikation gilt es die richtige Information, in geeigneter Form, zur richtigen Zeit am richtigen Ort verfügbar zu haben. Daraus ergeben sich für Unternehmen u.a. folgende Chancen, die zur Steigerung der Produktivität führen können: Gesteigerte Beratungsqualität, höherwertige Entscheidungsgrundlagen, effizientere Prozessabläufe und kürzere Entscheidungszeiträume.

Zusätzlich erfahren wir auch im privaten Bereich eine gesteigerte Lebensqualität. Diese äußert sich beispielsweise in neuen Kontaktmöglichkeiten zu unseren Mitmenschen (durch SMS, Email), ortsunabhängiger Verfügbarkeit von Daten (z. B. Nachrichten, Telefonnummern, Sportergebnisse), sowie in der Annahme von verbesserten Dienstleistungen. Aktuell bieten mobile, drahtlose Kommunikationsgeräte neue Möglichkeiten den bestehenden Informationsbedarf zu decken. Das Verhältnis von Kosten zu Nutzen verändert sich dabei zusehends zu Gunsten der Anwender dieser Technologie. Diese Tatsache in Verbindung mit einer Vielfalt an sich bietenden Einsatzmöglichkeiten sorgen für schnelle Verbreitung derartiger mobiler Kommunikations- und Informationssysteme. Die Ausbreitung neuer Technologien birgt jedoch zumeist auch neue Risiken. Die vorliegende Seminararbeit beschreibt eines der Risiken, welches beim Einsatz von drahtloser Netzwerktechnologie zum Tragen kommt: Die unberechtigte Nutzung und das Ausspionieren von Funknetzen.

Zugleich erleichtert diese Seminararbeit den Einstieg in die Thematik WLAN (Wireless Local Area Network).

Die Aktualität des Themas „War Driving" spiegelt sich u. a. in der Geschwindigkeit wieder mit der sich z. Zt. bestehende Internetforen in diesem Bereich entwickeln. Während der dreimonatigen Ausarbeitungszeit dieses Themas stieg die Anzahl der Beiträge im Deutschen War Driving Forum um ca. 300%.

Inhaltsverzeichnis

1.1 Einleitung 9
1.1.1 Vorgehensweise 9
1.1.2 Orientierung 10
1.2 Wireless LAN 10
1.2.1 Aufbau- und Funktionsüberblick 10
1.2.2 Der Standard IEEE 802.11 12
1.2.3 Aktuelle Entwicklung im Bereich WLAN 15
1.3 „War Driving“ - Begriffsklärung/Hintergründe 16
1.3.1 Entstehung des War Driving 16
1.3.2 Idee und Absicht 17
1.3.3 Interessengemeinschaften und Vereinigungen 18
1.4 Technische Ausrüstung 18
1.4.1 Benötigte Hardware 18
1.4.2 Benötigte Software 20
1.5 Einsatzbeispiele von War Driving 22
1.5.1 Ergebnisauswertung 22
1.5.2 Sicherheitsaspekte 23
1.5.3 Rechtliche Betrachtungsweise 24
1.6 Ausblick/Risiken 26

1.1 Einleitung

Sowohl in geschäftlichen als auch in privaten Bereichen sind Lokal Area Netzwerke (LAN) heute weit verbreitet. Während das Vorhandensein kabelgebundener Vernetzung mittlerweile weitgehend als selbstverständlich angenommen wird, rückt derzeit ein neuer Trend in den Mittelpunkt allgemeinen Interesses:

WLAN (Wireless Local Area Network)

Die aktuell geschaffenen technischen Voraussetzungen versprechen flexiblen Netzzugang, steigende Übertragungsraten und längere Nutzungszeiten aufgrund sparsamen Energieverbrauches. Umgesetzt werden die Technologien in Handys, Personal Digital Assistants (PDAs), Handhelds oder Notebooks und Laptops. Aufgrund der breitbandigen Anbindung sowie allgegenwärtiger Netzzugangsmöglichkeit („ubiquity of access“) erwachsen viel versprechende Anwendungsmöglichkeiten. Wie schon mehrfach innerhalb der IT-Branche beobachtet, vollziehen sich Evolutionsschritte gelegentlich in Form eines Hype (z. B. UMTS Hype, XML Hype, E-Business Hype). Kennzeichnend dafür sind anfänglich hohe z. T. unrealistische Erwartungen in eine neuartige Technologie, die anschließend in einen Interessenabfall und letztendlicher Konsolidierung in realistische Umsetzungsmöglichkeiten münden. Die Entwicklung von WLAN Systemen hat mittlerweile die Phase der Konsolidierung erreicht [1]. Ein großer Markt an WLAN Produkten existiert bereits und der Kampf um Marktführerschaft und Nischenplatzierung ist in vollem Gange. Jedoch birgt jede neue Technologie auch neue Gefahren in sich. Diese Seminararbeit soll dazu beitragen Chancen und Risiken im Bereich WLAN zu identifizieren und besser abwägen zu können.

1.1.1 Vorgehensweise

Absicht der vorliegenden Arbeit zum Thema „War Driving - An Approach to Locate Open WLANs“ ist es, Technologie und Ideologie dieses neuartigen Betätigungsfeldes darzulegen. Auf der Grundlage von Standards nach IEEE[1] soll im Kapitel 1.1.2 ein Überblick über aktuelle Entwicklungen im WLAN-Bereich vermittelt werden. Im darauf folgenden Abschnitt wird die Frage geklärt: „Was ist War Driving?“ Ausgehend von dieser Tätigkeitsbeschreibung werden anschließend im Kapitel 1.1.4 alle erforderlichen Ausrüstungsgegenstände und der Anforderungen an diese beschrieben. Der Umgang mit den genannten Ausrüstungsgegenständen wird in Kapitel 1.1.5 dargestellt. Dadurch soll ein Eindruck aus der Praxis von War Drivern vermittelt werden. Neben den Darstellungen aus der „Alltagstätigkeit“ von War Drivern wird in diesem Kapitel auch Sicherheitsaspekten sowie der rechtlichen Betrachtungsweise nachgegangen, um den Einblick in die War Driving Szene zu komplettieren.

[1] In der Kommunikations- und Datenbanktechnik existieren verschiedene Normierungsgremien. Ein wichtiges Gremium, welches neben dem OSI (Open Systems Interconnection) steht, ist das IEEE (Institute of Elektrical and Electronic Engineers) [2] mit Sitz in New York. Dieses Institut befasst sich mit der Standardisierung von elektronischen Systemen in verschiedenen Gebieten [3].

1.1.2 Orientierung

Diese Arbeit stellt die Tätigkeit einer Gruppe von Gleichgesinnten dar, deren Ziel es ist, mittels ausgefeilter Technik und zum Teil anspruchsvoller Hard- und Software Zugang zu drahtlosen Netzwerken zu erhalten. Eine Bewertung dieser Handlungsweise ist nicht Absicht der Arbeit. Vielmehr soll dem Leser durch Aufklärung über die aktuellen technischen Möglichkeiten, die Vorgehensweisen innerhalb der War Driving Communities, sowie anhand der Darstellung der gesetzlichen Hintergründe eine eigene Urteilsbildung ermöglicht werden. Im Fokus stehen dabei theoretische, technische und gesetzliche Grundlagen die Einfluss auf Übertragungsraten, Reichweiten und Frequenzauswahl haben, da diese Faktoren für den War Driver von maßgeblicher Bedeutung sind. Aufbauend auf diesen Grundlagenkenntnissen erfolgen Beschreibungen zum Aufbau und Einsatz von Hard- und Software für War Driver. Die dargestellten Einflussfaktoren und Grundprinzipien liefern die Rahmenbedinungen für erfolgreiches oder nicht erfolgreiches War Driving.

1.2 Wireless LAN

Im Folgenden Abschnitt werden Funktion, Aufbau und Standardisierungsdetails von Wireless LANs beschrieben. Ziel ist es, ein Grundverständnis für die technischen Voraussetzungen zu vermitteln. Diese Voraussetzungen bilden die Basis des Umfeldes in dem sich alle War Driving Tätigkeiten bewegen.

1.2.1 Aufbau- und Funktionsüberblick

Der Aufbau bzw. Ausbau eines WLANs (vgl. Abbildung 1.1) ist bequemer als der Aufbau bzw. Ausbau eines LANs:

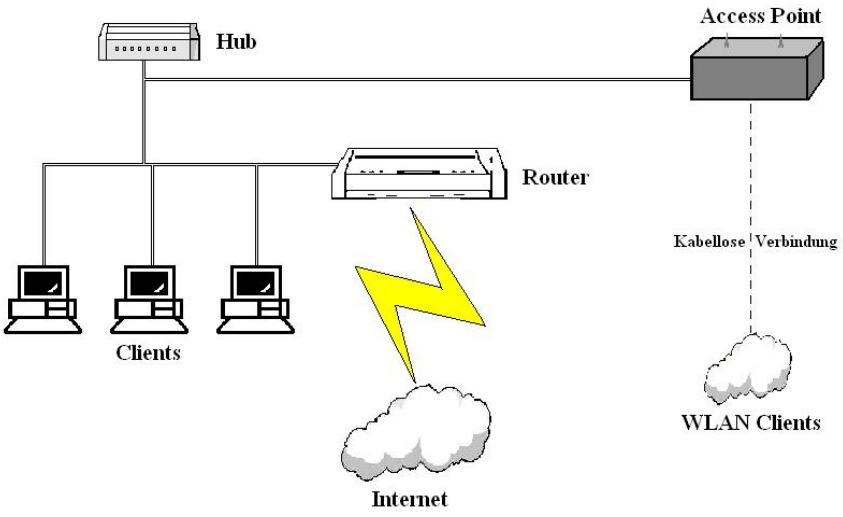

Abbildung 1.1: WLAN Architektur (`Infrastructure-Modus`) [5]

> *„Anstatt zum Beispiel in einem Firmengebäude in die Wände Löcher für die Kabel zu bohren, reicht es, eine zentrale Funkstation, Access Point, an das Strom- und lokale Netz anzuschließen, in den Rechner die Funknetzkarte einzubauen und die Software zu installieren. Dieser Access Point übernimmt dann die Versorgung mit der drahtlosen Netzanbindung und dient als eine Ethernet-Brücke.“* [4]

Während ein Access Point (AP) den Nutzer (client) immer mit genau einem Netzwerk verbindet, eignen sich WLAN Router zur Anbindung der Nutzer an mehrere Netze. Dabei übernimmt der Router die IP-Adresse des clients in seine Routingtabelle, während der AP sämtliche IP-Adressen ignoriert und alle Datenpakete weiterleitet [6].

Solange sich ein WLAN Nutzer in Empfangsreichweite zum Access Point aufhält kann Datentransfer in beiden Richtungen zwischen WLAN Nutzer und LAN stattfinden. Im Bezug auf die Empfangsreichweite ist folgendes zu beachten:

Die Ausbreitung von Funkwellen hängt neben der Sendeleistung von verschiedenen örtlichen Gegebenheiten ab, die Einfluss auf Absorption, Reflexion und Bündelung der Funkwellen haben. Für diese Arbeit besteht Relevanz nur insofern, als dass festgestellt werden kann, dass die Ausbreitung der Funkwellen eines APs oder eines WLAN Routers nicht auf die geographische Grenze eines Gebäudes oder eines Grundstückes begrenzt ist. Folglich ist der Netzzugang zu einem WLAN schwieriger zu kontrollieren als zu einem kabelgebundenen LAN.

Neben dem in Abbildung 1.1 dargestellten „Infrastructure Modus“ existiert noch eine zweite Architekturform, der „Ad Hoc Modus“. Sie wird an dieser Stelle der Vollständigkeit halber erwähnt. Eine Darstellung dieses Modus liefert die Abbildung 1.2.

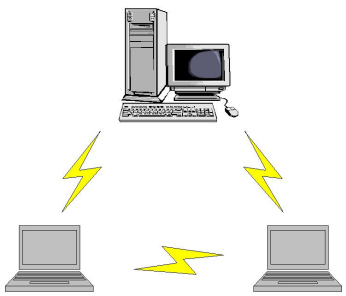

Abbildung 1.2: WLAN Architektur (`Ad Hoc Modus`) [7]

Für die weitere Betrachtung im Rahmen dieser Seminararbeit ist letztgenannte Architektur nicht von Bedeutung, da War Driver auf der Suche nach wieder auffindbaren, stationären APs sind. Aufgrund der Mobilität von Ad Hoc Netzen ist eine Auswertung und Informationsweitergabe über aufgefundene WLAN-Architekturen dieser Art nicht sinnvoll.

1.2.2 Der Standard IEEE 802.11

Um den Weg für die Entwicklung und Herstellung der ersten WLAN-Geräte zu ebnen begann das IEEE Ende der 90er Jahre mit der Ausarbeitung eines Standards nach dem WLAN-Geräte zu konstruieren seien, um Kompatibilität und Mindestfunktionsumfang sicherzustellen. Mittlerweile existiert unter der Bezeichnung IEEE 802.11... eine ganze Familie an Standards. Eine Übersicht zu allen Standards nach IEEE 802.11... ist in der Tabelle 1.1 dargestellt.

Tabelle 1.1: Übersicht zu den Standards nach IEEE 802.11... [8]:

IEEE-	**Beschreibung**
802.11	Protokoll und Übertragungsverfahren für drahtlose Netze, 1997 für 2 Mbit/s bei 2,4 GHz definiert
802.11a	54-Mbit/s-WLAN im 5-GHz-Band, 12 nicht-überlappende Kanäle, Modulation: Orthogonal Frequency Division Multiplexing (OFDM)
802.11b	11-Mbit/s-WLAN im 2,4-GHz-Band, 3 nicht-überlappende Kanäle
802.11c	Wireless Bridging zwischen AccessPoints
802.11d	„World Mode“, Anpassung an regionsspezifische Regulatorien (z. B. Frequenzbereich)
802.11e	QoS- und Streaming-Erweiterung für 802.11a/g/h (Priorisierung von Datenpaketen, z. B. für Multimedia-Anwendungen, Streaming)
802.11f	Roaming für 802.11a/g/h (Inter Access Point Protocol IAPP), bezogen auf AccessPoints verschiedener Hersteller
802.11g	54-Mbit/s-WLAN im 2,4-GHz-Band, Modulation OFDM
802.11h	54-Mbit/s-WLAN im 5-GHz-Band mit DFS[2] und TPC[3]
802.11i	Verbesserung von Authentifizierung/Verschlüsselung für 802.11a/b/g/h (AES, 802.1x) Ergänzend/Aufbauend auf WEP[4] u. WPA[5] (Interimslösung)
802.11j	Japanische Variante von 802.11a für den Bereich 4.9 GHz - 5 GHz
802.11k	Bessere Messung/Auswertung/Verwaltung der Funkparameter (z.B. Signal-stärke), soll z. B. Ortsbezogene Dienste (location-based services) ermöglichen
802.11ma	Zusammenfassung früherer Ergänzungen, Bereinigung von Fehlern aus vorausgegangenen Spezifikationen (Maintenance)
802.11n	Geplante Erweiterung für zukünftiges, schnelleres WLAN mit mindestens 100 Mbit/s

[2]Dynamic Frequency Selection

[3]Transmit Power Control

[4]Wired Equivalency Privacy

[5]WiFi Protected Access (WiFi = „wireless Fidelity“: Standard, um Kompatibilität zwischen „wireless“-Geräten unterschiedlicher Hersteller zu gewährleisten.)

Im Folgenden Abschnitt liegt der Fokus auf den Standards, die für das War Driving in Deutschland relevant sind. Diese werden ausführlich beschrieben. Zunächst sind ihre wichtigsten Eigenschaften in einer tabellarischen Übersicht (Tabelle 1.2) dargestellt:

Tabelle 1.2: Übersicht zu relevanten WLAN-Standards im Hinblick auf War Driving (Quellen: [9], [10] und [11]):

WLAN-Standards im Überblick				
Standards:	IEEE 802.11a	IEEE 802.11h	IEEE 802.11b	IEEE 802.11g
Bandbreite:	54 Mbit/s	54 Mbit/s	11 Mbit/s	54 Mbit/s
Frequenz:	5-GHz-Band	5-GHz-Band	2,4-GHz-Band	2,4-GHz-Band
Reichweite:	12 m - 20 m	12 m - 20 m	30 m - 300 m	20 m - 100 m
Besonderes:	weniger Störungen, höchste Kosten	TPC[6], DFS[7] (Europäische Variante von IEEE 802.11a)	Etabliert, niedrigste Kosten	Kompatibel zu IEEE 802.11b, Kosten: „IEEE 802.11b“+10%

IEEE 802.11 Der Standard IEEE 802.11 wurde Ende des Jahres 1999 als Norm für Wireless LANs verabschiedet [12]. Die Spezifikation beschreibt die drahtlose Schnittstelle *„over-the-air interface“* zwischen dem client und der Basisstation bzw. zwischen dem client und dem AP.

> *„Dieser Standard war der allererste drahtlose WLAN-Standard vom IEEE. Er ermöglichte die Produktion von Produkten mit einer Bruttoübertragungsrate von 1 Mbit/s. Er fand in dieser Form keine große Nutzergruppe.“* [13]

Der Standard IEEE 802.11 legt neben den Verfahren zur Authentisierung, nach *Identify-based-* oder *Challenge-Response*-Methode, noch den Aufbau des WEP-Protokolls (Wired Equivalency Privacy) sowie das WEP-Verschlüsselungsverfahren fest. Der WEP Schlüssel wurde vor ca. 2 Jahren gebrochen und ist deshalb in dieser Form unsicher [4].

IEEE 802.11a Dieser Standard setzt auf nutzbare Frequenzbereiche im 5-GHz-Band. Die vorhandenen 12 Kanäle sind nicht überlappend. Die Frequenzzuweisung für Europa ist folgendermaßen festgelegt: Die Bänder 5,15 bis 5,35 GHz und 5,47 bis 5,725 GHz stehen zur Verfügung. Über eine größere Anzahl an Subcarriern (als bei IEEE 802.11b) und OFDM-Modulation[8] wird eine Bruttodatenrate von maximal 54 Mbit/s erreicht. Alle Geräte nach diesem Standard müssen Übertragungsraten von 6, 12 und 24 Mbit/s unterstützen. Optional sind 9, 18 und 36 Mbit/s. Aufgrund der Möglichkeit des Mehrfachzugriffes durch aktive Nutzer eines WLANs entsteht auf der Ebene 3 (dem Network Layer) des OSI-Schichtenmodells zusätzlicher Verwaltungsaufwand. Dieser sorgt dafür,

[6] Transmit Power Control
[7] Dynamic Frequency Selection
[8] Orthogonal Frequency Division Multiplexing

dass im günstigsten Falle von 54 Mbit/s noch rund 32 Mbit/s zur Verfügung stehen [13]. Weiterhin sinkt der Nettodatendurchsatz mit steigendem Abstand zwischen Sender und Empfänger. Typischerweise verringert sich der Datendurchsatz in etwa um 300 Kbit/s pro Meter. Generell erreichen Sender dieses Standards eine Reichweite von 30 m bis 50 m. Die Reichweitenangaben differieren in den verschiedenen Literaturquellen, vermutlich aufgrund unterschiedlicher Messverfahren (sie liegen zwischen 12 m [10] und 50 m [13]).

Für den Einsatzbereich War Driving bedeutet dies, dass aufgrund der angegebenen Reichweite, ein geographisch ortsnaher Zugang zum AP erforderlich ist. Eine Möglichkeit das Reichweitenproblem zu verringern, besteht in dem Einsatz einer ausgefeilten Antenne oder Antennenanlage (vgl. Kapitel 1.4.1).

IEEE 802.11b Seit 2001 dominieren auf dem Markt WLAN-Geräte nach dem Standard IEEE 802.11b. Genutzt wird hier das 2,4-GHz-ISM[9]-Frequenzband. In Deutschland sind 13 Kanäle zur Nutzung freigegeben, von denen drei jeweils nicht überlappend angeordnet sind. Erreicht wird eine Datenübertragungsrate von 11 Mbit/s. Aufgrund der relativ schmalen Bandbreite und zusätzlicher Nutzung des 2,4-GHz-ISM-Frequenzbandes von Bluetooth- oder Microwellengeräten sind gegenseitige Beeinträchtigungen zu erwarten [13]. Die Funkwellen dieses Frequenzbandes erfahren aufgrund der, im Vergleich zum 5-GHz-Band, niedrigeren Sendefrequenzen eine geringere Dämpfung. Die mögliche Reichweite liegt daher jenseits der Reichweite von IEEE 802.11a-Geräten zwischen 30 m und 300 m [10].
Besonders stark verbreitet sind Geräte nach IEEE 802.11b in den USA [9]. Jedoch auch in Europa greifen WLAN Nutzer und damit ebenfalls War Driver auf Geräte dieses Standards zurück.
Als Konsequenz für den War Driver ergibt sich bei der Verwendung dieser Geräte, aufgrund des Reichweitenvorteils häufiger die Möglichkeit APs aufzuspüren, als mit einer Ausrüstung nach IEEE 802.11a. Zwar ist eine geringere Datenübertragungsrate hinzunehmen, jedoch sorgt der Kostenvorteil der IEEE 802.11b-Geräte für Akzeptanz und Verbreitung innerhalb der War Driving Communities.

IEEE 802.11g Um der Forderung nach mehr Bandbreite sowie Abwärtskompatibilität zum Standard IEEE 802.11b nachzukommen, wurde im Juli 2003 IEEE 802.11g verabschiedet [14]. Als obligatorisches Modulationsverfahren wurde OFDM aus IEEE 802.11a übernommen. Die Verträglichkeit aller Betriebsmodi zu IEEE 802.11b ist sichergestellt. Über den Standard IEEE 802.11g wird eine Datenübertragungsrate von 22 Mbit/s bis 54 Mbit/s ermöglicht. Die Reichweite im 2,4-GHz-Band liegt zwischen 30 m und 50 m. Der Standard stellt das Ende der Entwicklung in diesem Frequenzband dar. Es ist denkbar dass aufgrund der starken Nutzung dieses Frequenzbereiches und der damit verbundenen Störanfälligkeit in naher Zukunft auf das störungsfreiere 5-GHz-Band ausgewichen wird. Auch steht im 5-GHz-Band mehr Bandbreite zur Verfügung [13].
Für den Bereich War Driving in Deutschland zählen Geräte die der Norm IEEE 802.11g entsprechen vorerst zur meist genutzten Hardware.

[9]ISM Industrial Scientific Medical: Lizenzfreies Frequenzband in dem sich Bluetooth-, IEEE 802.11-, IEEE 802.11b- und Mikrowellengeräte bewegen.

IEEE 802.11h Bei diesem Standard handelt es sich um eine Erweiterung des IEEE 802.11a. Der Standard IEEE 802.11h ist abwärtskompatibel zu IEEE 802.11a. Ein Grossteil der Erweiterungen betrifft die interne Betriebssoftware. Ebenfalls im 5-GHz-Band bietet IEEE 802.11h eine Übertragungsrate von bis zu 54 Mbit/s bei gleicher Reichweite wie IEEE 802.11a. Mittels Implementierung von DFS und TPC stellt dieser Standard eine Art Endpunkt der WLAN-Verfahren in diesem Frequenzband dar. An Leistungsfähigkeit und Flexibilität bedeuten die beiden Standards IEEE 802.11g und IEEE 802.11h das Nonplusultra im jeweiligen Frequenzband. Dieser Standard liegt bisher in Form eines Entwurfes (draft) vor. Mit der Ratifizierung und anschließender Verfügbarkeit erster Produkte wird im Laufe des Jahres 2004 gerechnet [13].
Der Markt bietet derzeit viele Geräte nach IEEE 802.11a, IEEE 802.11b und IEEE 802.11g Standard. Beide letztgenannten sind unter War Drivern weit verbreitet. Inwieweit sich Geräte nach IEEE 802.11h im War Driving Bereich durchsetzten werden bleibt abzuwarten. Ebenso wie nach IEEE 802.11a wird der bestehende Reichweitennachteil möglicherweise durch geringere Störungsanfälligkeit ausgeglichen. Letztendlich werden sehr wahrscheinlich die Kosten, insbesondere im Vergleich zu Geräten nach IEEE 802.11g, über die zu erreichenden Marktanteile entscheiden.

IEEE 802.11i Hier handelt es sich um eine Erweiterung der Standards IEEE 802.11a, IEEE 802.11b sowie IEEE 802.11g, um bekannte Sicherheitslücken des WEP zu schließen. Beabsichtigt ist die Einführung eines neuen Protokolls. Mit Implementierung des Temporal Key Integrity Protocol (TKIP) kommen rotierende Schlüssel zum Einsatz, die jeweils nach kurzer Lebensdauer durch neue ersetzt werden. Der Standard existiert derzeit als draft und soll in Kürze ratifiziert werden [12], [15].
Für War Driver wird die Umsetzung des Standards eine neue Herausforderung darstellen. Galt die Aktivierung von WEP bisher nicht mehr als nennenswertes Hindernis beim Versuch Zugang zum AP zu erhalten, so verspricht IEEE 802.11i auch WEP2002 genannt vorerst hochgradige Sicherheit aufgrund der Verfahren[10] Re-Keying, Per-Packet-Mixing, Re-Sequencing und Message-Integrity-Check (MIC).

1.2.3 Aktuelle Entwicklung im Bereich WLAN

Derzeit nimmt die Verbreitung von WLANs trotz Sicherheitsbedenken zu. Es besteht bei vielen Unternehmen ein reges Interesse an drahtlosen Netzen. Durch den Einsatz drahtloser Technologie soll die Mobilität der Mitarbeiter und damit auch die Produktivität gesteigert werden. Da die Systeme einfach zu installieren sind und heute in aller Regel per Plug and Play in Betrieb genommen werden können besteht eine entsprechend große Nachfrage [16]. Nach Schätzung der Gartner-Group soll es 23 Millionen WLAN-Nutzer bis zum Jahre 2007 geben [17].
Weiterhin existiert die Forderung nach größerer Bandbreite, ähnlich der 100 Mbit/s in LANs. Dazu wurde durch das IEEE im September 2003 die Arbeitsgruppe IEEE 802.11n eingerichtet. Ziel ist die Erreichung einer Datenübertragungsrate in WLANs von 100 Mbit/s effektiv. Mögliche Techniken können sein Channel Bonding, Antennenarray (MIMO[11]).

[10] Die Darstellung der genannten Verfahren ist u.a. in [10] S. 280 f. zu finden.
[11] Models and Infrastructures for Mobile Computing

Dadurch sollen zukünftig neue Anwendungen insbesondere im Bereich Multimedia möglich werden. Die Fertigstellung des Standards wird für Oktober 2005 erwartet [18]. (Zwischenzeitliche Meldungen, dass der Standard früher verabschiedet werden würde sind dementiert.)
An einer anderen Entwicklung arbeitet derzeit das Europäische Institut für Telekommunikations-Standards (ETSI) [12]. Die Arbeit an einem Standard unter dem Namen HiperLAN/2[12] hat an dieser Stelle jedoch lediglich informativen Charakter. Für das War Driving besteht keine Relevanz, da es keinerlei Produkte nach diesem Standard gibt und die Zukunft dieses Standards ungewiss ist. Integrierte Schnittstellen zum Mobilfunk der dritten Generation (G3 oder UMTS) sowie zum High-Speed-Bus (IEEE1394) postulieren für die Zukunft ein „integratives Netzwerk“ [13] unter dem unterschiedlichste Dienste zusammengefasst werden können. Auch hier geht der Trend in Richtung zunehmender Mobilisierung, bei großer Bandbreite und steigenden Übertragungsgeschwindigkeiten unter Erfüllung bestehender Qualitätsanforderungen.

1.3 „War Driving“ - Begriffsklärung/Hintergründe

Für den ersten Teil des Begriffes kristallisiert sich in Insiderkreisen die Bedeutung Wireless Access Revolution (WAR) heraus. Der letztere Teil des zweiteiligen Begriffes „Driving“ bezeichnet das Umherfahren in Automobilen. Die Ausübung von War Driving kann jedoch auch zu Fuß erfolgen und auch das War Cycling (Radfahren) wird erwähnt. Aus verschiedenen Definitionen, die im Internet zu finden sind, hier zwei der treffendsten:

> War Driving: *„Die Suche nach drahtlosen Netzwerken durch Einwahl (auf dem Fußweg oder während des Fahrens (in Pkw oder Bus)) mittels drahtlosem Einwahlgerät. Gelegentlich unterstützt durch Verwendung von hochverstärkenden Antennen und GPS-Geräten (Global Positioning System).“* [19]

> War Driving: *„Das Umherfahren (im Pkw) bei gleichzeitiger Suche nach ungeschützten drahtlosen Netzwerken - Begriffsprägung durch Peter Shipley.“* [19]

1.3.1 Entstehung des War Driving

Peter Shipley beschäftigte sich um die Jahrtausendwende intensiv mit der Sicherheit von Netzwerken. Zunächst war sein Blickpunkt ausgerichtet auf die Sicherheit von Netzwerkverbindungen über kabelgebundene Modems. Eine seiner Veröffentlichungen befasst sich mit der „Analyse von Einwahlmodems und deren Sicherheitslücken“ [20]. Ausgangspunkt war der Hollywood Film „War Games“ (1983). Der Film lieferte breiten Anlass für Computer- und Netzwerkinteressierte zu Erproben, inwieweit sie durch Wählen zufälliger Telefonnummern Zugang zu ungesicherten Netzen erhalten konnten. Damals wurde der Begriff „War Dialing“ geprägt [21].

[12] High Performance Radio Local Area Network, Type 2

Eine parallele Entwicklung vollzog sich aufgrund der erweiterten Bandbreite die mit der Einführung von LANs (Local Area Network) geschaffen wurde. Der Versuch eigene ungenutzte Bandbreite für andere zugänglich und nutzbar zu machen (vgl. Kapitel 1.3.3), fand schnell Übertragung in den WLAN-Bereich. Anfänglich wurden öffentlich zugängliche APs mit Tafelkreide auf Bürgersteigen oder an Hausfassaden markiert (Warchalking) [22]. Später gingen Menschen auf die Suche nach bisher unveröffentlichten Zugangsknoten. Maßgeblich beteiligt an der Prägung des Begriffes „War Driving“ war Peter Shipley [19]. Zwar fuhren andere vor Ihm durch Stadt und Land auf der Suche nach ungeschützten kabellosen Zugangsknoten (wireless access point (AP)). Jedoch war Peter Shipley der erste, der die Suche und Sammlung von Informationen über Zugangsknoten komplett automatisierte. Die Automatisierung erfolgte unter Zuhilfenahme eines GPS-Gerätes und ausgewählter Software (vgl. Kapitel 1.4.2).

1.3.2 Idee und Absicht

Es stellt sich die Frage: Aus welchem Grund investiert jemand Zeit, Geld und Arbeit in eine Tätigkeit, für die die Aussichten auf Dank oder Entlohnung als gering anzusehen sind?

Im Rahmen der Arbeit an diesem Thema fanden sich keinerlei empirische Untersuchungen/Untersuchungsergebnisse, die gesicherte Erkenntnisse über die Intentionen der War Driver liefern könnten. Den Internetforen ist zu entnehmen, dass die Motivation zum War Driving sehr differenziert sein kann:
Sie mag von wissenschaftlichem Interesse in Bezug auf Ausbreitung und Nutzung der WLAN-Technologie und/oder dem praktischen Testen von Sicherheitsmechanismen, über den Wunsch nach kostenlosem Netzzugang, oder dem Reiz des Neuen, Elitären, Halblegalen bis hin zu dem Drang sich unberechtigt Daten und Informationen zu verschaffen (mit destruktivem Charakter) gehen.
Letzteres gehört eindeutig nicht zu den Intentionen der War Driver Gemeinschaft (community). Erkenntlich wird das daran, dass auf allen Internetseiten zu War Driving sowie auf allen Veröffentlichungen zu War Driving folgender Hinweis zu finden ist:

> *„`Reminder`. Don't break the LAW while wardriving, this will only give us negative feedback and will hurt everyone in wardriving community.“* [19]

Die Art der Veröffentlichungen lässt darauf schließen, dass der eigentliche Zweck des War Driving die Sammlung von Informationen ist. Diese werden aufaggregiert und in Form von Kartenmaterial (vgl. Kapitel 1.5.1) oder Tabellen veröffentlicht. Auf diese Weise wird auf Sicherheitslücken hingewiesen und zur Sensibilisierung wenig sicherheitsbewusster WLAN-Nutzer beigetragen. Außerdem werden gewonnene Erfahrungen und Informationen im Rahmen von War Driving Treffen ausgetauscht und ergänzt.

1.3.3 Interessengemeinschaften und Vereinigungen

Interessengruppen bilden sich z. Zt. vielerorts über das Internet. Während sich in den USA bereits in vielen Städten WAR Driving Gemeinschaften (communities) formiert haben, gibt es Bundesweit erst eine einzige. Auf der Internetseite „Deutsches Wardriving-Portal“ [23] wird die Anzahl der registrierten Benutzer mit über 1.000 angegeben (Stand: Jan. 2004). Aktuell formieren sich War Driving Communities in den großen Städten wie Berlin, Düsseldorf, Frankfurt, Hamburg, München und weiterer.
Auch in anderen Ländern Europas entstehen Internetportale und Foren zum Thema War Driving (vgl. Tabelle 1.3 im Anhang A). Die Internetseite „www.Mobileaccess.de“ [24] ermöglicht die Suche nach vorhandenen öffentlichen Hot-Spots in Deutschland, Österreich und der Schweiz. Der starke Zuwachs an Informationen sowohl auf dieser Internetseite als auch in verschiedenen War Driving Foren kann als Nachweis dienen, dass Interesse und Mitgliederzahlen im War Driver Umfeld in diesen Monaten eine besonders große Wachstumsrate erleben.
Während die Zusammenarbeit in Europa noch im Aufbau ist, gibt es in den USA bereits regelmäßige Treffen und intensiven Informationsaustausch [25]. Einige Gruppen verfolgen heute die Zielsetzung ein für jedermann frei zugängliches WLAN für den geographischen Bereich ganzer Städte einzurichten - so genannte FreeNets (vgl. Tabelle 1.4 im Anhang B). Die Vorstellung geht dahin, dass private Haushalte ihren Zugang zum AP offen lassen. So können andere clients diesen Zugang nutzen, während man selber, quasi im Gegenzug, andere WLAN-Verbindungen nutzt/ nutzen kann sobald man sich außer Haus begibt.

1.4 Technische Ausrüstung

Das Angebot an technisch geeigneter Ausrüstung für den War Driver ist vielfältig. Sowohl im Hardwarebereich als auch im Bereich der Software existiert bereits ein breiter Markt an Produkten. Die Darstellung einer umfassenden Übersicht würde den Rahmen dieser Arbeit sprengen. Daher wird im Folgenden das Hauptaugenmerk auf häufig genutzte Hard- und Softwareprodukte gelegt.

1.4.1 Benötigte Hardware

Grundsätzlich eignet sich jede mobile Hardware mit der Zugang zu WLAN (wireless access) hergestellt werden kann. Im Rahmen von Durchführung/Ausübung des War Driving kommen Notebooks, Laptops oder PDAs zum Einsatz. Diese werden um ein Schnittstellengerät zum Funkempfang (wireless device), eine Antenne und ggf. um ein GPS-Gerät ergänzt (vgl. Abbildung 1.3).

Notebook[13]/PDA Um der Anforderung nach Eignung für Netzwerktechnologie gerecht zu werden, sollte das Gerät über eine CPU mit einer Taktfrequenz von 400 MHz (oder höher) verfügen. Die sinnvolle Größe für den Arbeitsspeicher liegt bei 128 MB RAM.

[13]Notebook wird hier als Synonym für Notebooks und Laptops eingesetzt.

Sofern diese Hardware die Ausführung eines Betriebssystems (vgl. Kapitel 1.4.2) unterstützt sind alle erforderlichen Voraussetzungen gegeben [26].

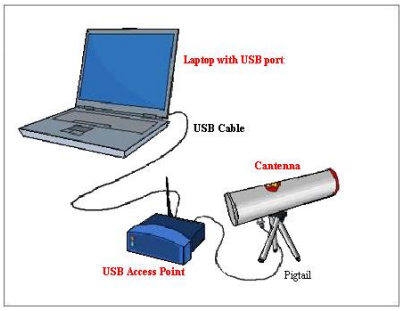

Abbildung 1.3: War Driving System [19]

Wireless Device Die derzeit verfügbaren Wireless LAN Karten unterstützen den Anschluss an eine von zwei möglichen Schnittstellen. Es existieren PCMCIA-Karten und USB-Karten. Für beide Arten sind sowohl Ausführungen nach IEEE 802.11a als auch nach IEEE 802.11g erhältlich. Ein Vorteil der USB-Ausführung liegt in der Möglichkeit das Gerät flexibel im Fahrzeuginnenraum zu platzieren. Nicht alle Karten bieten die Möglichkeit zum Anschluss einer Antenne. Um War Driving Resultate zu optimieren ist dies ein wichtiges Kriterium [23].

Antennen Bilder und Diskussionen aus bestehenden Internetforen machen deutlich, dass dem Einfallsreichtum bei der Auswahl einer Antenne kaum Grenzen gesetzt sind. Schon ein Eigenbau [27] ermöglicht eine Verstärkung des Eingangssignales um bis zu 12 db. Dadurch ergibt sich für War Driver beim Aufspüren von Netzen ein Antennenvorteil mit einem Faktor von 8 [28].
Es kommen unterschiedliche Antennentypen zum Einsatz (Beispiele: Stabantennen, Drahtantennen, Hohlstäbe, Miniaturantennen). Generell lässt sich zwar die Aussage vertreten *„je größer die Antenne desto besser der Empfang“* [27]. Jedoch muss die Antenne im Pkw handhabbar bzw. eine sichere Montage gewährleistet sein.
Obwohl die Empfängerempfindlichkeit gesetzlich nicht begrenzt ist, gilt auch für den Einsatz von Antennen, dass nicht alles was von War Drivern eingesetzt werden kann den gesetzlichen Bestimmungen entspricht.

GPS-Einheit Zur eigentlichen Ausübung des War Driving wird ein GPS-Gerät nicht unmittelbar benötigt. Es erleichtert jedoch das Auffinden bereits bekannter APs und ist darüber hinaus erforderlich, um APs in eine geographische Kartensoftware einzutragen. (Beispielsweise: Microsoft Mappoint oder Microsoft Autoroute Express [26].) Auch der GPS-Markt bietet eine große Varietät an Geräten. Für den Einsatz im Rahmen von War Driving Fahrten sind keinerlei besondere Anforderungen erforderlich. Ein Gerät welches dem aktuellen Stand der Technik entspricht reicht daher für War Driving Zwecke aus.

Spannungsversorgung Zum Betrieb eines einzelnen Notebooks oder PDAs ist die Anschaffung eines geeigneten 12V-Adapters ausreichend. Für den gleichzeitigen Betrieb mehrerer Einheiten im Rahmen von Gruppen War Driving im gleichen Fahrzeug, geht aus den Internetforen die Empfehlung hervor einen Spannungswandler einzusetzen. Diese liefern 110 bzw. 220 Volt und leisten Unterstützung für maximal drei Geräte.

1.4.2 Benötigte Software

War Driving Software gibt es für nahezu alle Betriebssysteme (z. B.: Windows, Linux, Mac-OS, BSD-OS (Berkeley Software Design-Operating System), WinCE, WinMobile 2003). Einschränkungen bestehen bei Win98 und Früheren aufgrund mangelhafter Netzwerkfunktionalität. Die gesamte Software im Hinblick auf das War Driving Tätigkeitsfeld umfasst Anwendungsprogramme zum Aufspüren von wireless APs, Scannen und Überwachen von Netzwerken sowie zur Erstellung von Landkarten. Treiber und Hilfsprogramme zur Anbindung der individuellen Hardware seien hier nur der Vollständigkeit halber erwähnt.

Ermitteln von Access-Standorten Zum Aufspüren von APs eignen sich Netstumbler und ISS unter Windows, Wellenreiter und Kismet unter Linux sowie Kismac und Macstumbler unter Macintosh. Bei allen Produkten handelt es sich um Freeware. Aufgrund der steigenden Nachfrage an derartigen Werkzeugen und dem aktiven Einsatz durch War Driver haben die aufgeführten Programme bereits einen qualitätssichernden Entwicklungsprozess durchlaufen. Im Folgenden soll ein kurzer Einblick in die Möglichkeiten und Anwendungsweise derartiger Software am Beispiel von Netstumbler gegeben werden:

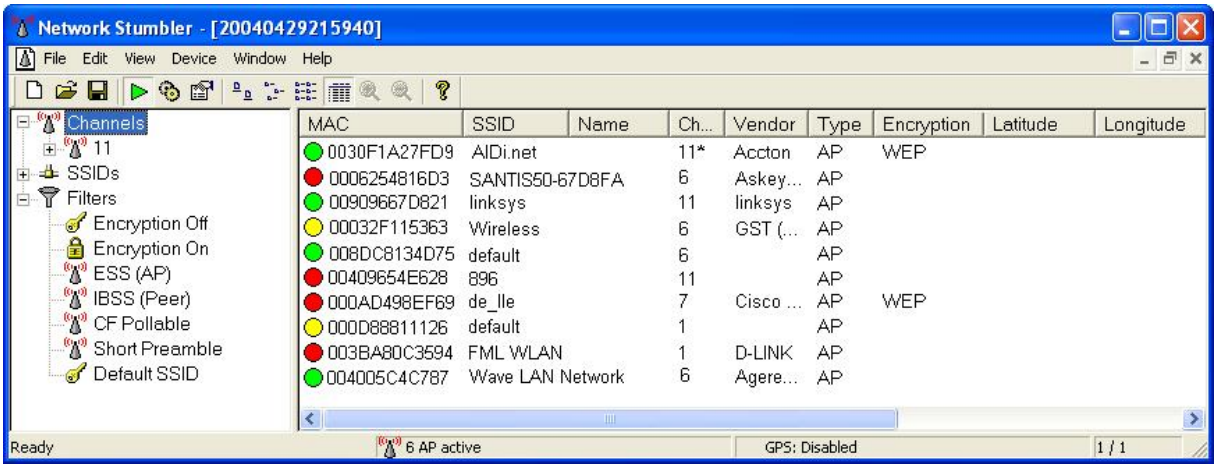

Abbildung 1.4: Netstumbler: Access point found [23]

Im Anwendungsfenster des Netstumblerprogrammes (vgl. Abbildung 1.4) wird jeder gefundene Zugangsknoten aufgelistet. Es werden u. a. Daten zu folgenden Kategorien ermittelt und aufgelistet: Empfangsanzeige, MAC-Adresse, SSID, Sende- und Empfangskanal, Hersteller des AP (Erlaubt Rückschlüsse auf Standardpasswörter), Typ (AP), Signalstärke, Verschlüsselungsinformationen, GPS-Daten. Den Farben der Empfangsanzeige kommt jeweils folgende Bedeutung zu:

`Grün`: Guter Empfang `Gelb`: Mäßiger Empfang `Rot`: Kein ausreichender Empfang

Zusätzlich besteht die Möglichkeit die eingehende Signalstärke anzeigen zu lassen. Die Darstellung erfolgt im Programm in graphischer Form (vgl. Abbildung 1.5) und ermöglicht die Auswahl eines APs mit ausreichender Sendeleistung.

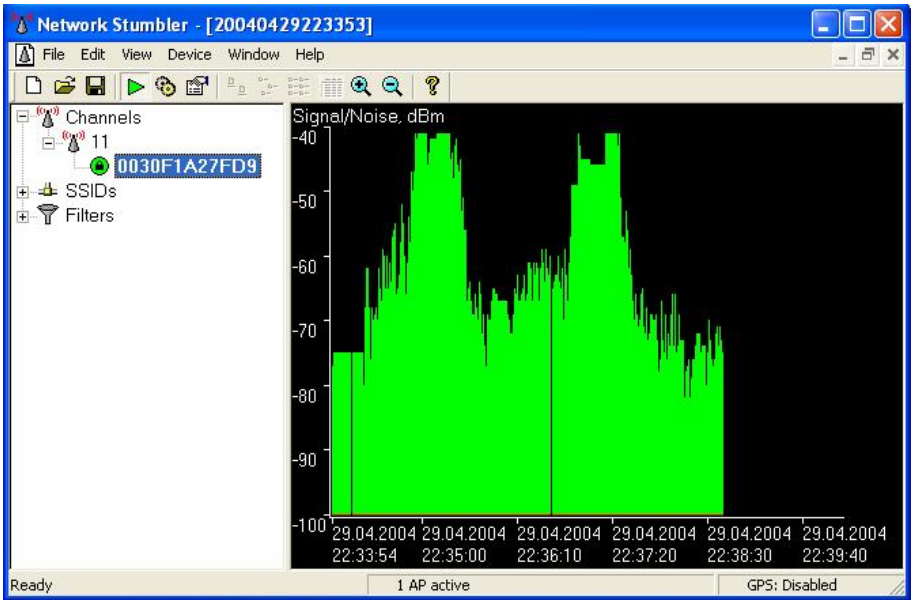

Abbildung 1.5: Netstumbler: Modus-Sendeleistung [23]

Die generell einfache Anwendung erlaubt auch nicht versierten Computernutzern schnelles Auffinden von APs und damit den leichten Einstieg in das War Driving.

IDS oder Sniffer Softwareprodukte dieser Kategorie ermöglichen auf unterschiedliche Art und Weise das Eindringen in Netzwerke. Unter Umgehung von Sicherheitsmaßnahmen können beispielsweise Netzwerkdaten ausspioniert werden. Daher ist diese Art von Software als sicherheitsgefährdend einzustufen. Eine große Auswahl an Programmen unterschiedlicher Funktionalität wird im Internet dokumentiert oder zum download angeboten. Fertige Softwarelösungen existieren nicht nur zur Ermittlung des WEP-Schlüssels (vgl. Kapitel 1.2.2 (S. 9), Kapitel 1.5.2 und [29]). Auch das komfortable Abhören (vgl. Aerosol (PRG) [30]) und Analysieren von Datenpaketen (vgl. Colasoft Capsa (PRG) [31]) und sogar das Vortäuschen eines AP mittels einer einfachen WLAN-Netzwerkkarte ist möglich (vgl. HostAP (PRG) [32]). Die aufgeführten Tools sollen als Beispiele dienen welche Möglichkeiten sich dem War Driver bieten [29]. Diese kleine Auswahl macht deutlich, welche Chancen und Risiken in derartiger Software liegen. Chancen bieten sich für Administratoren und authorisiertem Personal zur Prüfung ihres Netzes. Risiken bestehen im feindlichen Einsatz solcher Werkzeuge.

Allgemeine Ergänzungen Nicht jede Software ist kompatibel zur Hardware. Die Internetforen (vgl. [23], [25] und Anhang A) liefern eine Fülle an Informationen über mögliche erfolgreiche Systemkonfigurationen. Zu vielen Problemen bestehen bereits Erfahrungswerte auf die angehende War Driver zurückgreifen können.
Ein interessanter Softwarebereich aus War Driving Sicht ist das Angebot an Kartensoftware. Programme wie Stumbverter (Windows) oder GPSD (Linux) komplettieren das Softwarepaket eines War Drivers. Mit ihnen lassen sich punktgenau geographische Positionen innerhalb digitaler Landkarten eintragen und speichern. Dazu werden über eine Schnittstelle geographische Positionen z. B. aus der Netstumbleranwendung entnommen und in vorhandene digitale Kartendateien exportiert [19].
Es klang bereits an, dass während der Entwicklung von Softwareprodukten im Bereich

drahtloser Netzwerktechnik die rechtlichen Bestimmungen der verschiedenen Länder, in denen die Tools zum Einsatz kommen, keine Berücksichtigung finden. Leicht gerät ein Nutzer durch Installation und Anwendung eines Freeware Tools in den Bereich der Illegalität. Die Grenzen sind dabei fließend. Aufgrund unterschiedlicher Funktionalität innerhalb der Anwendungsprogramme lässt sich eine klare Trennung zwischen legaler und illegaler Software oft nicht treffen.
Es hängt zuletzt zum Einen vom Verantwortungsbewusstsein des Nutzers und zum Anderen von der Errichtung ausreichender Sicherheitsvorkehrungen durch den Betreiber ab, inwieweit rechtliche Bestimmungen eingehalten werden.

1.5 Einsatzbeispiele von War Driving

In diesem Abschnitt soll zunächst ein Eindruck aus der Praxis vermittelt werden. Anhand eines Kartenausschnitts wird aufgezeigt in welcher Art die Darstellung und Auswertung gewonnener Informationen nach einer War Driving Einsatzfahrt erfolgen kann. Im Anschluss folgen Überlegungen zu Sicherheitsaspekten und rechtlichen Hintergründen.

1.5.1 Ergebnisauswertung

Die Abbildung 1.6 zeigt das Ergebnis einer War Driving Fahrt vom 14. Januar 2004 aus Lancaster (USA):

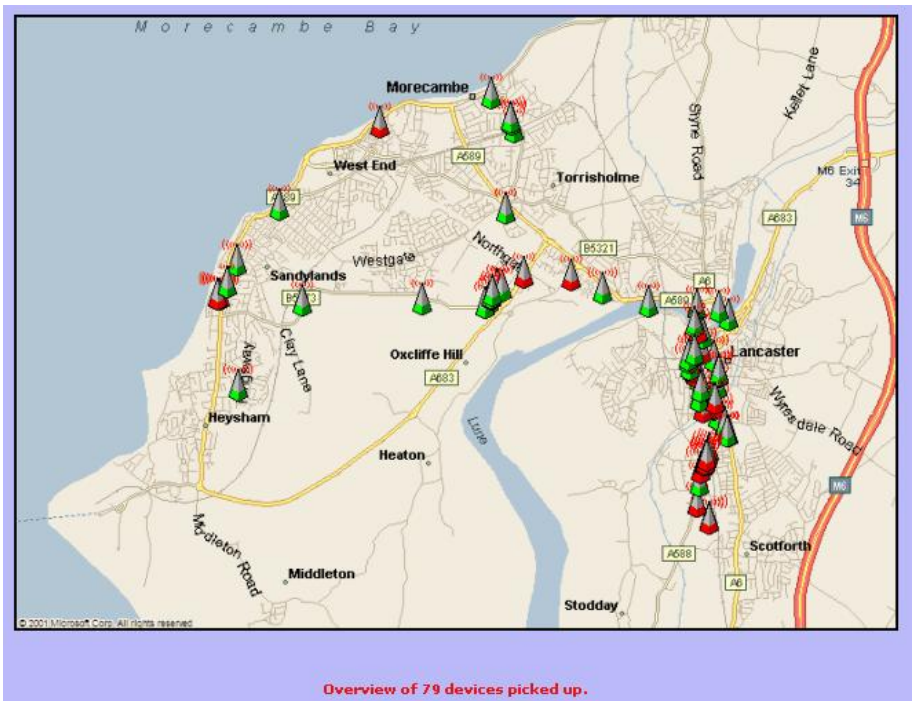

Abbildung 1.6: Resultat eines War Driving Einsatzes [19]

Während dieser Fahrt wurden 79 APs ermittelt. Sämtliche Zugangsdaten wurden gespeichert. Mittels GPS wurden gleichzeitig die Symbole für Zugangspunkte durch die Software in den entsprechenden Positionen auf der Karte eingetragen.
Durch WEP geschützte Zugangsknoten werden dabei als AP-Symbol in roter Farbe vermerkt (). Ungeschützte APs entsprechen den grünen Markierungen ().

Wird die selbe Route mehrfach abgefahren, können verlässliche Informationen gewonnen werden. Anhand der gespeicherten Daten sind dann statistische Aussagen zu relevanten Fragestellungen möglich, wie z. B:

- Wie vollzieht sich die Ausbreitung von WLAN Systemen?
- Wie schnell wächst die Anzahl an WLANs?
- Wie groß ist der Anteil an ungesicherten APs?
- Steigt der Anteil an gesicherten APs?
- Gibt es lokale Unterschiede bezüglich des Sicherheitsbewusstseins?

So lassen sich Aussagen treffen inwieweit Aufklärungsarbeit von War Drivern oder Informationssendungen in TV und Rundfunk das Verhalten der WLAN-Administratoren beeinflusst. Über den Vergleich bzw. Austausch der Ergebnisse mit anderen War Drivern sind zusätzlich auch Aussagen über die Qualität der jeweils genutzten War Driving Ausrüstung möglich.

1.5.2 Sicherheitsaspekte

An dieser Stelle wird nicht auf die Gefahren während der Ausübung von War Driving, die sich vorwiegend auf sicheres Führen des Kraftfahrzeuges beschränken, eingegangen. Vielmehr sollen sicherheitsrelevante Aspekte während des Betreibens von WLANs aufgezeigt werden die sich z. T. aus den Inhalten des Kapitels 1.2.2 ergeben:

Bei der Einrichtung von Hot-Spots[14] in Form von Bürgernetzen (Flughäfen, Bahnhöfe, Cafe, Messen, etc.) ist zu beachten, dass von Seiten der Klienten kein Zugriff auf sicherheitsrelevante Daten möglich ist. Daher sollten die APs keine direkte Anbindung an das lokale Netzwerk besitzen, sondern zuvor eine Firewall überbrücken müssen [9].
Ebenso besteht für den Einsatz von drahtlosen Firmennetzen ein, im Vergleich zum LAN, erhöhtes Sicherheitsrisiko. Aufgrund der Eigenschaften von Funkwellen (vgl. Kapitel 1.2.1) ist die Zugangskontrolle mit erhöhtem softwaretechnischen Aufwand verbunden. Möglichkeiten bieten **1)** die Aktivierung von WEP, **2)** die Zulassung oder Auswahl bestimmter MAC[15]-Adressen, **3)** Unterbindung der Übertragung des Service Set Identifiers (SSID) oder **4)** die Einrichtung eines Virtual Private Network (VPN).

[14] Geographischer Ort, an dem drahtloser breitbandiger Internetzugang über einen Zugangsknoten (AP) zur Verfügung gestellt wird.
[15] Media Access Control

1) Problematisch bei WEP ist, dass der RC4-Schlüssel mittels Brute Force Verfahren durch heutige Softwareprodukte in relativ kurzem Zeitraum errechnet werden kann. Entsprechende Tools [16] (Wepcrack, Airsnort, Kassandra) sind im Internet erhältlich und fester Bestandteil im Werkzeugkoffer von War Drivern.

2) Die Zugangssicherung mittels Angabe von MAC-Adressen erfordert manuelle Verwaltungsarbeit, die sich nur bei kleineren Netzen noch in angemessener Zeit bewältigen lässt. Diese Schutzmaßnahme lässt sich durch MAC-Spoofing[16] umgehen. Auch hierzu wird per Internet in den War Driving Foren unterstützende Software angeboten.

3) Ein Basic Service Set (BSS) kann durch einen Netzwerknamen, den SSID identifiziert werden. Das Unterdrücken der Übertragung des SSID ist per Default-Einstellung zunächst nicht aktiviert. Sobald es aktiviert ist besteht ein gewisser Schutz gegen Eindringversuche einiger WLAN-Scanner. Jedoch ermitteln verschiedene andere frei erhältliche Tools die SSID dennoch anhand von Steuer und Managementsignalen.

4) Größtmögliche Sicherheit bietet die Einrichtung eines VPN, welches zwischen den Clients und der lokalen Firewall implementiert wird. Zusätzlich sollten ständige Kontrollen durch ein Wireless Intrusion Detection System erfolgen.

Die dargelegten Informationen zu möglichen Sicherheitsmaßnahmen machen deutlich, dass mit Ausnahme von **4)** Errichtung eines VPNs, jede einzelne Maßnahme nur unzureichenden Schutz gegen Angriffe auf WLANs bietet. Da die Maßnahmen zu den Punkten **1)** bis **3)** mittlerweile leicht zu umgehen sind, bietet erst wieder die Kombination mehrerer Sicherheitsmaßnahmen einen wirkungsvollen Schutz gegen Datenraub und -missbrauch. Dabei sollte unternehmensweit eine ganzheitliche Sicherheitspolicy aufgesetzt werden. Für regelmäßige „Risk Assessments“ [9] bietet sich die Nutzung eines Dienstleisters in IT-Risikofragen an. Generell ist eine Aufwand-Nutzenrechnung zu erstellen, auf deren Basis eine Abschätzung über den Umfang sinnvoller Sicherheitsmaßnahmen erfolgen kann.
In vielen Fällen werden aufgrund von Unwissenheit die verfügbaren Sicherheitsverfahren nicht eingesetzt. Die Tatsache, dass in den Defaulteinstellungen zur Einrichtung von wireless LANs alle Sicherheitsmaßnahmen deaktiviert sind trägt nicht zur Aufklärung bzw. zur Steigerung der Netzwerksicherheit bei. Aber selbst wenn implementierte Maßnahmen eingesetzt werden, bieten sie häufig nicht den gewünschten Schutz. In der aktuellen Situation unterstützen Softwaretools aus dem Internet den War Driver bei der Umgehung der meisten Sicherheitsmaßnahmen. Erst die Implementierung zukünftiger neuer Sicherheitstechniken z. B. nach dem Standard IEEE 802.11i (vgl. Kapitel 1.2.2) bietet im Bereich WLAN erhöhte Sicherheit.

1.5.3 Rechtliche Betrachtungsweise

Der §3 des Telekommunikationsgesetzes (TKG) vom 25. Juli 1996 beinhaltet eine gesetzliche Begriffsbestimmung zu den Worten „Betreiben“, „Endeinrichtungen“, „Funkanlagen“,

[16] Vorspiegelung einer gefälschten MAC-Adresse.

„Netzzugang“, „Nutzer“ und Weiterer. Aus Sicht des War Driving ist neben diesen Grundlagen insbesondere der §86 Satz 2 TKG von Interesse. Dort heißt es:

> *„Mit einer Funkanlage dürfen Nachrichten, die für die Funkanlage nicht bestimmt sind, nicht abgehört werden. Der Inhalt solcher Nachrichten sowie die Tatsache ihres Empfangs dürfen, auch wenn der Empfang unbeabsichtigt geschieht, auch von Personen, für die eine Pflicht zur Geheimhaltung nicht schon nach § 85 besteht, anderen nicht mitgeteilt werden.“*

Neben dem deutschen Telekommunikationsgesetz kommen auch europaweite Regelungen zum tragen: Mit In-Kraft-Treten der „Frequenzentscheidung“[17] der EU zum 25. Juli 2003 gilt folgende Regelung:

> *„Funk-LAN-Systeme dürfen entweder das [...] 2,4-GHz-Band*[18] *[...] oder die [...] 5-GHz-Bänder*[19] *[...] ganz oder teilweise*[20] *nutzen.“ [33]*

Bis dato war nicht klar, ob freie offene Bürgernetze, in Form grundstücks-übergreifender Funknetze unter die Lizenzpflicht nach Paragraph 6 des Telekommunikationsgesetzes (TKG) fallen. Diese Unsicherheit ist nun behoben und es heißt „freies Funken“.
Zusammenfassend lässt sich festhalten: Das Senden und Empfangen im lizenzfreien Frequenzspektrum der genannten 2,4-GHz- und 5-GHz-Bänder unterliegt keinen Einschränkungen.
Bezüglich des Zugangs zu Netzwerken sieht es ein wenig anders aus. Sobald ein Nutzer online geschaltet ist werden Datenpakete empfangen. Nach dem Auszug aus §86 Satz 2 TKG dürfen nur solche Daten empfangen werden, die auch an diesen Nutzer gerichtet sind. Nun können War Driver durchaus Daten aus dem Internet beziehen, die für sie bestimmt sind (z. B. private email). Hier beschreiten War Driver eine rechtliche Grauzone. Zum Einen mögen die Inhalte durchaus an sie selbst gerichtet sein, zum Anderen nutzen sie fremde Netzwerke und empfangen Pakete von Sendeanlagen, die nicht für ihre persönliche Nutzung in Betrieb genommen wurden. Es steht dabei außer Frage, dass die Anwendung einiger Funktionen vorhandener Tools, die das Ausspionieren von Netzen, aufbrechen von Sicherheitsschlüsseln, vorspiegeln falscher Tatsachen (Mac-Spoofing) ermöglichen, als illegal einzustufen ist. Diesbezüglich wird in den erwähnten War Driving Internetforen an jede bereitgestellte Information oder Anwendungssoftware der Hinweis gekoppelt diese, im Sinne des Ehrenkodex (vgl. Kapitel 1.3.2), nicht für den Gesetzesmissbrauch zu nutzen. Nach §95 TKG droht demjenigen Freiheitsstrafe bis zu zwei Jahren oder Geldstrafe, der entgegen §86 Satz 1 oder Satz 2 eine Nachricht abhört oder den Inhalt einer Nachricht oder die Tatsache ihres Empfanges einem anderen mitteilt. An dieser Stelle bleibt zu diskutieren inwieweit War Driving als ehrenhafte positive Aufklärungsarbeit anzusehen ist, oder ob der Ehrendcodex der War Driving Vereinigungen (vgl. Kapitel 1.3.2) mentale Rechtfertigung und Motivation liefert, um gesetzliche Grenzen zu überschreiten.

[17] Entscheidung Nr. 676/2002/EG des Europäischen Parlaments und des Rates vom 7. März 2002 über einen Rechtsrahmen für die Funkfrequenzpolitik in der Europäischen Gemeinschaft.

[18] Frequenzband von 2400,0-2483,5 MHz.

[19] vgl. Kapitel 1.2.2

[20] Hier bezieht sich 'teilweise' auf den oberen Bereich des 5-GHz-Bandes, der in einigen Ländern für Radar benutzt wird.

1.6 Ausblick/Risiken

Vor der Ratifizierung des Standards IEEE 802.11g im Juli 2003 haben Anbieter wie 3Com, Cisco, Enterasys und Siemens keine so genannten Vorstandardprodukte (Pre-g-Produkte) auf den Markt gebracht [9]. Daher kommen derzeit (seit Abschluss der Ratifizierung) verstärkt IEEE 802.11g-Geräte auf den Markt. Dabei werden steigende Produktionszahlen dieser Hardware zu sinkenden Preisen auf dem Absatzmarkt führen. Darüber hinaus besteht weiterhin Nachfrage an zusätzlicher Bandbreite wie sie mittels IEEE 802.11n verwirklicht werden soll. Es ist folglich davon auszugehen, dass sich aufgrund der günstigeren Preise für WLAN-Technik, der zunehmenden Verbreitung von WLANs und mangelhaftem Sicherheitsbewußtsein vieler Nutzer die Betätigungsmöglichkeiten für War Driver ausweiten. Es bleibt zu hoffen, dass die steigende Zahl der War Driving-Community-Mitglieder sich an den selbst vorgegebenen Ehrenkodex (vgl. Kapitel 1.3.2) halten wird. Inwieweit bei statistisch steigenden Mitgliederzahlen sich auch die Anzahl „schwarzer Schafe“ erhöht ist nicht vorherzusehen. Für Nutzer von WLANs ergibt sich in jedem Falle ein Anstieg an Aufklärungsbedarf. Ob das Wissen um die Gefahren beim Einsatz von WLAN und die Bereitstellung von Werkzeugen zur Abschirmung (z. B. nach IEEE 802.11i) oder der Einfallsreichtum und die Finesse der War Driver schneller wächst bleibt abzuwarten. Da *„ein drahtloser Internetzugang bald genauso selbstverständlich wie die Strom- und Wasser-Versorung“* [10] sein wird, sollten Sicherheit und Authentizität von Daten m. E. auch in Zukunft auf drei Säulen aufbauen:

1. Verpflichtung zum gewissenhaften Umgang mit Software und Daten (nach Ehrenkodex (vgl. Kap.1.3.2) bzw. anhand des Normen- und Wertegefüges unserer Gesellschaft).
2. Zunehmende Aufklärungsarbeit der Anbieter und Hersteller.
3. Schaffung eindeutiger und durchsetzbarer Regelungen durch den Gesetzgeber.

Auf Basis eines starken Normen- und Wertegefüges in Verbindung mit wirkungsvollen Sicherheitsmechanismen wäre die zukünftige sinnvolle Gestaltung eines sicheren, global verfügbaren drahtlosen Internetzugangs, nach den Beispielen aus Anhang B, denkbar und wünschenswert. Die Erfahrungen und Untersuchungsergebnisse von War Drivern könnten auf dem Weg dorthin wertvolle Erkenntnisse über Sicherheitslücken und Abschirmungsunzulänglichkeiten liefern.

Literaturverzeichnis

[1] G. Oscar: Editorial: *Hype Cycle*, Funkschau, Nr. 19, Coburg, Germany, Sep. 2003

[2] IEEE: `http://standards.ieee.org`, 14. Jan. 2004

[3] KeyNet AG: *Das IEEE und seine Aufgaben - [...]*, `http://www.key-net.ch/techsite/pdffiles/AufgabenIEEE.pdf`, 14. Jan. 2004

[4] Mohamed Kallel: *Security in Wireless Networks*, B. Stiller, O. Braun, A. Heursch, Mobile Systems I, München, Germany, Dez. 2002

[5] Gulli.com: Wardriving FAQ: `http://www.gulli.com/dokumente/faq/wlan.html#2`, 14. Jan. 2004

[6] J. Geier: *Understanding Wireless LAN Routers*, `http://www.wi-fiplanet.com/tutorials/article.php/1586861`, 01. Feb. 2004

[7] Ad-hoc *!!!!*

[8] IEEE Org: `http://standards.ieee.org/cgi-bin/status?wireless`, 23. März 2004

[9] C. Kartes: *Die Zukunft von WLAN heißt 802.11g*, Funkschau, Nr. 10, Coburg, Germany, Mai 2003

[10] H. J. Rauscher: *Großbaustelle Wireless*, Funkschau, Nr. 22, Coburg, Germany, Nov. 2003

[11] E. Bloam: *Drahtlos Messen und Testen*, Funkschau, Nr. 23, Coburg, Germany, Nov. 2003

[12] F.-J. Kauffels: *Wireless LANs*, Bonn, Germany, 2002

[13] A. Kral, H. Kreft: *Wireless LANs Networker's Guide*, München, Germany, Apr. 2003

[14] Status of Projekt IEEE 802.11g: `http://grouper.ieee.org/groups/802/11/Reports/tgg_update.htm`, 27. Mrz. 2004

[15] Status of Projekt IEEE 802.11i: `http://grouper.ieee.org/groups/802/11/Reports/tgi_update.htm`, 27. Mrz. 2004

[16] C. Kartes: *WLAN-Sicherheit heute und morgen*, Funkschau, Nr. 25-26, Coburg, Germany, Dez. 2002

[17] M. Gongolsky: *Der drahtlose LANsinn*, http://www.spiegel.de/netzwelt/technologie/0,1518,241860,00.html, Spiegel Online, 29. Mrz. 2004

[18] Wi-Fi Networking News: http://wifinetnews.com/archives/002783.html, 14. Jan. 2004

[19] Wardriving FAQ: *What is Wardriving?*, http://www.wardriving.info/live/index.php, 20. Jan. 2004

[20] P. Shipley, S. L. Garfinkel: *An Analysis of Dial-Up Modems and Vulnerabilities(2001)*, http://www.dis.org/filez/Wardial_ShipleyGarfinkel.pdf, 26. Jan. 2000

[21] Martin Puaschitz: *WLAN - WarDriving [...]*, http://www.it-academy.cc/content/article_browse.php?ID=593#6, 30. Jan. 2004

[22] Warchalking: http://www.warchalking.org, 28. Jan. 2004

[23] Deutsches Wardriving-Portal: http://www.wardriving-forum.de/wiki/, 24. Jan. 2004

[24] Mobileaccess.de: *Wireless LAN HotSpots - in deiner nähe*, http://neu.mobileaccess.de/wlan/index.html?sid=, 24. Apr. 2004

[25] Wardriving: *Liste Amerikanischer Organisationen*, http://www.wardriving.com/links.php, 01. Feb. 2004

[26] God de Vader: *Wardriving*, http://users.skynet.be/multibox, 10. Feb. 2004

[27] R. Flickenger: *Antenna on the Cheap(er, Chip)*, http://www.oreillynet.com/cs/weblog/view/wlg/448, 30. Mrz. 2004

[28] A. von Obert: *WLAN-Reichweite erhöhen* http://www.techwriter.de/thema/wlan-rei.htm, 30. Mrz. 2004

[29] A. Hagedorn: *IEEE 802.11i Sicherheit in drahtlosen lokalen Netzen*, Diplomarbeit, TU Darmstadt, Nov. 2003

[30] Aerosol (PRG): *Crack-Tool zum Mitschneiden von Datenpaketen*, http://www.stolenshoes.net/sniph/aerosol.html, 30. Mrz. 2004

[31] Colasoft Capsa (PRG): *Analyse-Tool zur Netzwerküberwachung*, http://www.colasoft.com/products/capsa/index.php?id=30916w, 30. Mrz. 2004

[32] HostAP (PRG): *Crack-Tool zur Emulation eines APs mittels einer WLAN Netzwerkkarte*, http://hostap.epitest.fi/, 31. Mrz. 2004

[33] Freifunk.wiki: http://www.freifunk.net/wiki/GesetzeUndVorschriftenDieWLANBetreffen, 24. Mrz. 2004

A.1 War Driving Foren in Europa

Die Tabelle 1.3 vermittelt eine Übersicht über die Länder in denen War Driver bereits in Gemeinschaften organisiert sind. Einige der aufgelisteten Internetauftritte sind noch neu und z. T. in Ausarbeitung. Es ist anzunehmen, dass in Kürze weitere Länder hinzu kommen.

Tabelle 1.3: Internetadressen europäischer War Driving Foren (Stand: Mrz. 2004)

Land:	**Internetadresse:**
Bundesrepublik	http://www.wardriving-forum.de/wiki, 31. Jan. 2004
Belgien	http://www.wardrivers.be/phpBB2, 29. Mrz. 2004
Frankreich	http://craiefiti.free.fr, 31. Jan. 2004
Niederlande	http://forum.wirelessnederland.nl/viewtopic.php?t=1724, 31. Jan. 2004
Österreich	http://www.wgv.at, 31. Jan. 2004
Portugal	http://www.canariaswireless.net/modules.php?name=News&file=article&sid=169, 31. Jan. 2004
Schweiz	http://www.wardriving.ch, 31. Jan. 2004

A.2 Flächendeckender Zugang

Verschiedene Interessensverbände von War Drivern oder auch WLAN-Interessierte haben sich die Aufgabe gestellt flächendeckend freien Zugang in/zu WLANs zu ermöglichen (FreeNets). Das Prinzip, welches dabei zu Grunde gelegt wird ist das folgende: Nutzer die Providergebühren bezahlen, stellen Ihre ungenutzte Bandbreite zur Verfügung, so dass Passanten den bereitgestellten Onlinezugang nutzen können. Im Gegenzug erwarten die einzelnen Nutzer, dass sie sich ebenfalls kostenlos an anderen geographischen Orten (z. B. in der Stadt) in das Internet einwählen können. Auf diese Art („Geben und Nehmen“) soll ein flächendeckendes Gesamtnetz entstehen über das sich Bürger mobil und ortsunabhängig einwählen können.
Vorreiter dieser Vorhaben sind im Folgenden tabellarisch aufgelistet:

Name der Gemeinschaft:	**Internetadresse:**
NoCatNet	http://nocat.net, 01. Feb. 2004
NoVAWireless	http://www.cawnet.org, 01. Feb. 2004
NYCwireless	http://www.nycwireless.net, 01. Feb. 2004
PersonalTelco	http://www.personaltelco.net/static/index.html, 01. Feb. 2004
Seattle Wireless	http://www.seattlewireless.net/index.cgi/FrontPage?action=show&redirect=StartSeite, 01. Feb. 2004
SFLan Wireless Data	hhttp://www.archive.org/web/sflan.php, 01. Feb. 2004

Tabelle 1.4: Internetadressen zur Errichtung von FreeNets (Stand: Feb. 2004)

A.3 Verzeichnis der Abkürzungen und Akronyme

AP	Access Point (Zugangsknoten für schnurlose Datenübertragung)
BSD	Berkeley Software Design (basierend auf UNIX)
GPS	Global Positioning System
DFS	Dynamic Frequency Selection (Dynamische Frequenzwahl)
DHCP	Dynamic Host Configuration Protocol
ETSI	Europäisches Institut für Telekommunikations-Standards
HiperLAN/2	High Performance Radio Local Area Network, Type 2
IEEE	Institute of Electrical and Electronics Engineers („I-triple-E“)
ISO	International Standardisation Organisation
LAN	Local Area Network
MAC	Media Access Control
MIC	Message-Integrity-Check
MIMO	Models and Infra-structures for Mobile Computing
OFDM	Orthogonal Frequency Division Multiplexing
OS	Operating System (Betriebssystem)
OSI	Open Systems Interconnection
PDA	Personal Digital Assistant
RAM	Random Access Memory
SSID	Service Set Identifier
TKG	Telekommunikationsgesetzes
TPC	Transmit Power Control
VPN	Virtual Private Network
WEP	Wired Equivalency Privacy
WiFi	Wireless Fidelity (Kompatibilitätssiegel/Gütesiegel der Industrie)
WLAN	Wireless Local Area Network

Kapitel 2

Neue VPN-Lösungen

Atnarong Kongnin

Das Internet hat sich in letzten Jahren als universales und globales Kommunikationsmedium etabliert. Man verwendet dieses riesige weltweite Netzwerk zur Kommunikation, um verschiedene verteilte Standorte und Personen miteinander zu verbinden. Heutzutage ist die Sicherheit im Internet ein wichtiges Thema, da die Datenkommunikation im Internet nicht vertraulich ist. Die private Datenkommunikation erfordert jedoch die Vertraulichkeit, Integrität und Authentizität der Informationen. Diese Anforderungen kann man mit der Technik der virtuellen privaten Netzwerke erfüllen. In dieser Arbeit werden zunächst allgemeine Definition des VPNs erläutert, sowie die typische VPN-Szenarien, die man überall findet. Anschliessend werden die Anforderungen an VPNs vorgestellt. Danach wird die Sicherheit von Virtuellen Privaten Netzwerken anhand deren Authentifizierung und Verschlüsselungsverfahren erläutert. Zudem werden verschiedene Algorithmen beschrieben. Anschliessend werden die breit verwendeten VPN-Protokolle anhand deren Zuordnung zu dem ISO/OSI-Referenzmodell beschrieben, sowie andere Implementierungen der VPN-Protokolle. Im Fazit wird eine Zusammenfassung gebildet und einen Vergleich zwischen IPSec und SSL-VPN.

Inhaltsverzeichnis

2.1 Einleitung 33
2.2 Virtual Private Networks 34
2.2.1 Definition 34
2.2.2 VPN-Szenarien 35
2.2.3 Anforderungen an VPNs 36
2.3 Die Sicherheit Virtueller Private Netze 37
2.3.1 **Authentifizierung** 38
2.3.2 Verschlüsselungsverfahren 39
2.4 Zuordnung der VPN-Protokolle zu den 7 Schichten des ISO/OSI-Referenzmodelles 43
2.5 Zusammenfassung 50

2.1 Einleitung

VPN-Virtuelle Private Netzwerke stellen eine sichere Verbindung dar. Sie bieten Vertraulichkeit, Integrität und Authentizität der Informationen. Von großen Unternehmen, die möglicherweise zwischen verschiedenen Standorten kommunizieren bis zur privaten Kommunikation zwischen zwei Rechnern ist VPN die beste Lösung, um eine sichere Kommunikation und kostengünstige Verbindung herzustellen. Heutzutage werden mehrere VPN-Lösungen angeboten. Sie stützen sich auf unterschiedlichen Protokolle ab.

Da das Internet grundsätzlich ein offenes, ungeschütztes Netz ist, besteht die Gefahr, daß übertragene Nachrichten von Unbefugten abgehört oder manipuliert werden. Ein Beispiel ist eine Abhörstation der Nato Security Agency (NSA) in Bad Aibling, die 50 km. südlich von München liegt.(s. Abb.2.1). NSA, der geheimste Geheimdienst der USA hört in Deutschland alle Telefongespräche, Faxe und E-Mails ab. Die Aufgabe der NSA ist, jede Art von ausländischer Kommunikation, die für die Sicherheit der Nato von Interesse ist, abzuhören und zu decodieren. Die NSA spielt im Internet eine wichtige Rolle, da die NSA darauf drängt, daß die Internet-Software und die Web-Browser nur mit Verschlüsselungsverfahren arbeiten, die von den Cray-Superrechnern der NSA mit vertretbarem Aufwand dechiffriert werden können.[5]

Abbildung 2.1: Die US-Abhörstation in Bad Aibling [5]

Wegen der Bedrohungen im Internet werden immer neue VPN-Lösungen vorgestellt, wie IPSec-VPN oder SSL-VPN. SSL-VPN ist die neuste VPN-Lösung, die von der Industrie vorgestellt wird. Es bietet eine einfachere Benutzung als jede andere VPN-Lösung. Allerdings hat jede VPN-Lösung Vor- und Nachteile. Der Benutzer muss entscheiden, welche Lösung für seinen Zweck am besten geeignet ist.

2.2 Virtual Private Networks

2.2.1 Definition

VPN steht für “Virtual-Private-Network“ oder “virtuelles privates Netzwerk“.

Virtual bedeutet, daß es kein physikalisches Netz ist. Die Verbindungen existieren nur wenn sie benötigt werden.[11]

Private bedeutet, daß die Kommunikation vertrauenswürdig und nicht öffentlich durchgeführt wird.[3]

Eine allgemeine Definition, die für alle Arten von VPNs gilt, ist “Ein VPN ist ein Netzwerk, das ein öffentliches Netzwerk benutzt, um private Daten zu transportieren“[2]. Es gibt zwei Möglichkeiten, um private Daten über das öffentliche Netzwerk zu transportieren(siehe Abbildung 2.2). Die erste Möglichkeit sind die so genannten Mietleitungen oder Standardfestverbindungen, die an große Netzwerke oder Firmen zur ausschließlichen Nutzung vermietet werden. Obwohl diese teure Verbindung nur vom Mieter derselben benutzt werden, gibt es aber die Möglichkeit, daß die Kommunikation von dem Netzbetreiber abgehört wird. Eine andere Möglichkeit ist die Benutzung von VPN. Zur Sicherung der Daten wird die Kommunikation über das öffentliche Netzwerk verschlüsselt gesendet. Diese Technik wird Tunneling genannt. Man kann mit der Tunneling-Technologie Pakete eines Netzwerkprotokolls in Pakete eines anderen Netzwerkprotokolls kapseln und übers Netzwerk übertragen. Das VPN kann teure Mietleitungen zwischen verschiedenen Standorten durch virtuelle Verbindungen ersetzen. Da VPN bei Bedarf aktiviert werden kann, sparen die Firmen Kosten für ungenutzte Kapazitäten.

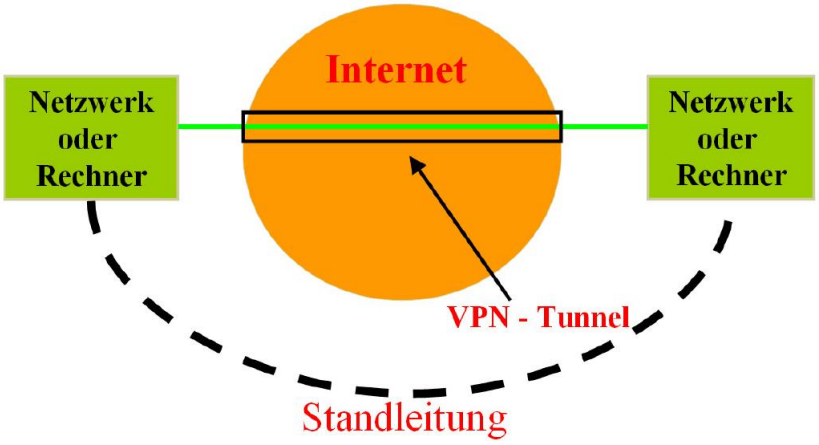

Abbildung 2.2: Virtual Private Networks

2.2.2 VPN-Szenarien

Es gibt heutzutage zahllose unterschiedlichen Arten von VPNs, mit spezifischen technischen Voraussetzungen. Man kann aber die verschiedene Arten von VPNs in 3 Szenarien, unterteilen.[7]

Host-to-Host

Diese Szenario stellt die sicherste Kommunikation zwischen zwei Rechnern über das Internet dar. Die ganze Verbindung von Host zu Host wird von dem Tunnel mit den verschlüsselten Daten abgedeckt. Eine Angriff auf der ganzen Verbindungslänge kann ausgeschlossen werden. Voraussetzung ist aber, dass jeder Host an der Kommunikationsschnittstelle mit entsprechender VPN-Software ausgestattet sein muss.

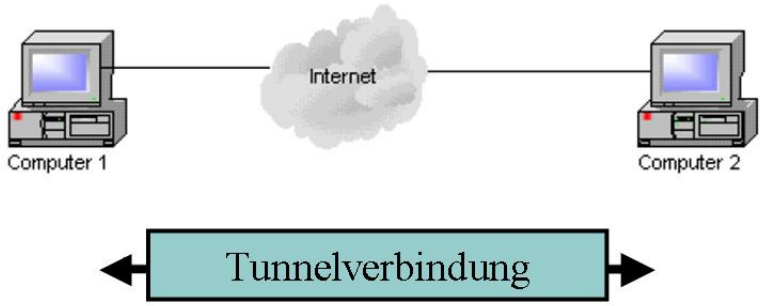

Abbildung 2.3: VPN-Szenarien 1 [6]

Gateway-to-Gateway

Bei diesem Szenario werden zwei VPN-Gateways zur Netzwerksicherung vor den Intranetzen positioniert. Die Kommunikation innerhalb eines Tunnels über das Internet ist verschlüsselt. Bei dieser Verbindung muss keine der lokalen Arbeitsstationen mit spezieller VPN-Software ausgestattet sein. Die ganze Arbeit betreffend die Sicherheit erledigen die Gateways, deshalb ist das VPN für die Rechner im Netzwerk vollständig transparent und senkt dieses Szenario den Verwaltungsaufwand für den Administrator durch ein VPN erheblich. Der Nachteil dieser Art der Verbindung ist, dass das Gateway sehr leistungsfähig sein muss, um alle Verbindungen zu ver- bzw. entschlüsseln.

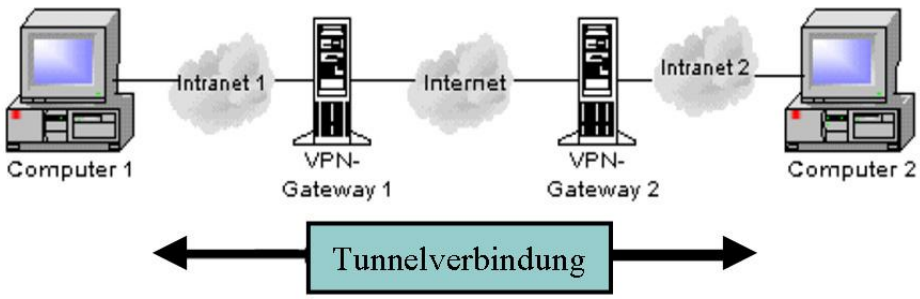

Abbildung 2.4: VPN-Szenarien 2 [6]

Gateway-to-Host

Bei der Gateway-to-Host Kommunikation werden die mobilen Clients bzw. Tele-Heimarbeiter ins VPN miteinbezogen. Dazu muss jeder Rechner mit spezieller VPN-Software ausgerüstet sein. Ein typisches Einsatzgebiet ist Remote Access über unsichere Transportnetze. Eine strenge Authentisierung hat eine große Bedeutung, um die Identität der mobilen Clients bzw Tele-Heimarbeiter genau zu überprüfen.

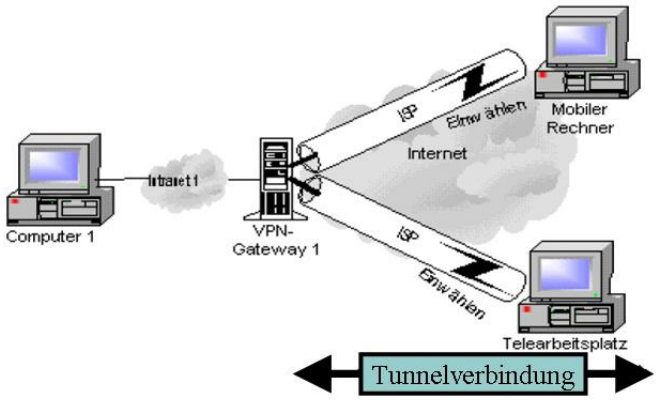

Abbildung 2.5: VPN-Szenarien 3 [6]

2.2.3 Anforderungen an VPNs

Wegen der verschiedenen Einsatzgebiete der virtuellen privaten Netzwerke, existieren viele verschiedene Anforderungen an VPNs wie z.B. Anforderungen an Sicherheit, Quality-of-Service, Verfügbarkeit, Performance sowie anderen Rahmenbedingungen, aber in dieser Seminararbeit möchte ich nur die Anforderungen an VPNs betreffend die Datensicherheit betrachten. Es gibt viele Ansprüche im Bereich der Datensicherheit, die sich jedoch in folgende Aspekte einteilen [2] :

- Datenvertraulichkeit
- Paket-Authentifizierung
- Datenintegrität
- Benutzer-Authentifizierung

1. **Datenvertraulichkeit**

   Eine sehr wichtige Anforderung an VPN ist die Datenvertraulichkeit. Die Daten auf dem Weg durch das Internet dürfen nicht von Unbefugten gelesen werden.

   Diese Anforderung erreicht man durch eine Verschlüsselung der Informationen durch den Absender. Durch die Entschlüsselung beim Empfänger werden die Informationen wieder lesbar.

2. **Paket-Authentifizierung**

   Paket-Authentifizierung garantiert, dass die Nachrichten oder ankommende Pakete vom richtigen Absender kommen und nicht von dritten mit gefälschten Absenderadressen und neu berechneten Prüfsummen geschickt wurden.

   Jedes ankommende Paket muss ähnlich wie bei einer Benutzer-Authentifizierung authentifiziert werden. Die Vertraulichkeit der Daten, die nur dem Sender und dem Empfänger bekannt sind, kann man durch symmetrische Schlüssel oder Pre-Shared Secrets erreichen.

3. **Datenintegrität**

   Ein weiterer Aspekt der Sicherheit ist die Datenintegrität. Der Empfänger muss erkennen können, ob ein ankommendes Paket während des Transport verändert wurde. Dazu wird die Paketprüfsumme mit speziellen Verfahren auf Basis von symmetrischen Verschlüsselungsverfahren berechnet.

   Die berechnete Paketprüfsumme wird in das Paket eingefügt. Die Schlüssel sind nur dem Sender und dem Empfänger bekannt. Wer ein Paket ändern will, kann die Prüfsumme nicht korrekt berechnen.

4. **Benutzer-Authentifizierung**

   Benutzer-Authentifizierung ist eine sehr wichtige Anforderung bei Remote-Access-VPNs. Eine Person, die über ein VPN-Gateway auf das Intranet zugreift, muss seine Identität nachgewiesen haben.

   Zahlreiche Authentisierungsmethoden können für die Bedürfnisse der verschiedenen VPN-Implementationen eingesetzt werden, z.B. die Benutzung von Passwort, Chipkarte oder die Überprüfung durch biometrische Verfahren wie Fingerabdruck-Leser. Heute kann man sowohl Authentifizierungsprotokolle wie z.B. PAP (Password Authentication Protocol) oder CHAP(Challenge Handshake Authentication Protocol) als auch Digitale Zertifikate zur Authentifizierung einsetzen, wie z.B. Digitale Zertifikate nach ITU-X.509-Standard.

2.3 Die Sicherheit Virtueller Private Netze

Die zu übertragenden Daten sollen vor verschiedenen Angriffen, wie z.B Denial-of-Service oder Replay-Angriffen geschützt werden. Die Sicherheit eines VPN kann durch die Authentifizierung und die Verschlüsselung auf verschiedenen Schichten des ISO/OSI-Schichtenmodells realisiert werden.

2.3.1 Authentifizierung

Man kann die Authentifizierung in zwei verschiedene Klassen unterteilen [2]:

- **Die Authentifizierung einer natürlichen Person**

  Hier wird die Identität einer nicht persönlich anwesenden Person geprüft. Wenn beispielsweise ein Benutzer auf ein Unternehmensnetzwerk über ein Remote-Access-VPN Zugriff haben möchte, muss er einem Zugangs-Kontrollsystem nachweisen, dass er tatsächlich der Berechtigte ist.

- **Die Authentifizierung von Übertragungseinheiten eines Netzwerkprotokolls, um bestimmte Angriffe wie Spoofing oder Man-in-the-Middle abzuwehren**

  Der Absender (ein VPN-Gateway oder ein Router) eines Pakets soll identifiziert werden. Diese Authentifizierung lässt sich in 2 Phasen aufbauen. Die erste Phase ist der Verbindungsaufbau, wobei der Absender und der Empfänger sich gegenseitig identifizieren. Die zweite Phase ist die Übertragungsphase, hier werden die Pakete überprüft, ob sie wirklich alle vom selben Absender kommen.

Authentifizierungssysteme und -protokolle

Die folgenden Verfahren und Technologien werden in verschiedenen Authentifizierungssystemen eingesetzt.

- **Password Authentication Protocol (PAP)**

  PAP ist ein standardisiertes Protokoll für Point-to-Point-(PPP)-Verbindungen. Das PAP ist ein einfaches Protokoll zum Austausch von Passwörtern. Die User-ID und das Passwort werden im Klartext zu einem zentralen Server übertragen, deswegen bietet das PAP-Protokoll nur geringen Schutz.[8]

- **Challenge Handshake Authentication Protocol (CHAP)**

  CHAP ist auch ein standardisiertes Protokoll für PPP-Verbindungen wie PAP und es ist eine Drei-Phasen-Handshake-Prozedur, in der das Kennwort nicht explizit wie beim PAP übertragen wird. Dabei überträgt der Server einen zufällig generierten Schlüssel zum Anwender. Das Passwort des Anwenders wird mit diesem Schlüssel verschlüsselt und es wird an den Server zurückgesendet. Der Server entschlüsselt nun das Passwort mit seinem Schlüssel und überprüft es.[9]

- **RADIUS**

  RADIUS (Remote Authentication Dial in User Service) ist ein Protokoll zur Authentifizierung von Remote-Nutzern. Es arbeitet mit einer Client-Server-Architektur. RADIUS stellt Mechanismen zur Benutzeridentifizierung über PAP und CHAP, zur Zugriffskontrolle über eine eigene RADIUS-Datenbank und zur Verwaltung von dynamischen IP-Adressen bereit. Die Passwörter werden im Klartext gespeichert, aber es gibt auch RADIUS-Server, die Passwörter verschlüsseln und wieder entschlüsseln können.

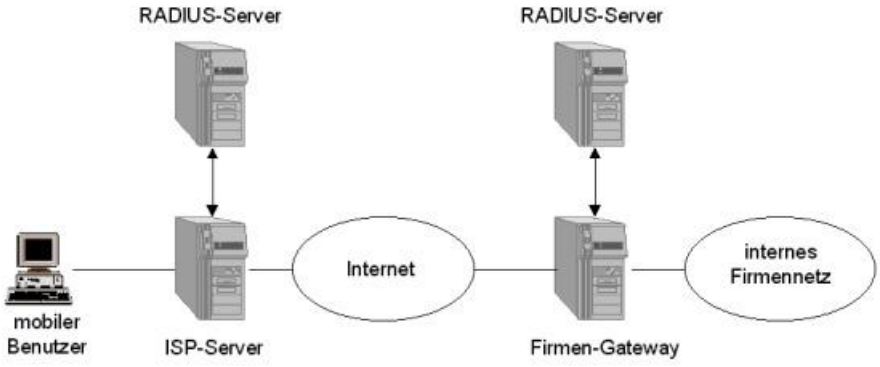

Abbildung 2.6: Aufbau eines VPN mit RADIUS-Servern[10]

2.3.2 Verschlüsselungsverfahren

Es gibt zwei grundsätzliche Verschlüsselungsverfahren, die auf Schlüsseln basieren. Das erste, die symmetrische oder Secret-Key-Verschlüsselung. Das andere ist die asymmetrische oder Public-Key-Verschlüsselung.[2]

1. **Symmetrische Verschlüsselungsverfahren**

   Bei einem symmetrischen Verschlüsselungsverfahren kennen alle beteiligten Gegenstellen nur einen geheimen Schlüssel, den man zur Verschlüsselung und zur Entschlüsselung verwendet. Da es nur einen, geheimen Schlüssel gibt, spricht man bei Kommunikation mit symmetrischer Verschlüsselung oft von "private key communication" . Das Grundprinzip funktioniert, wie in Abb.2.7 dargestellt. Zunächst muß der Schlüssel über einen sicheren Kanal ausgetauscht werden. Dieser Schlüssel dient dem Sender zur Verschlüsselung des Klartexts und dem Empfänger zur Entschlüsselung des Chiffretexts. Die Teilnehmer können danach die Nachrichten, die vom Sender gesendet werden, entschlüsseln. Selbst kann der Sender die Nachrichten für den Empfänger verschlüsseln. [12]

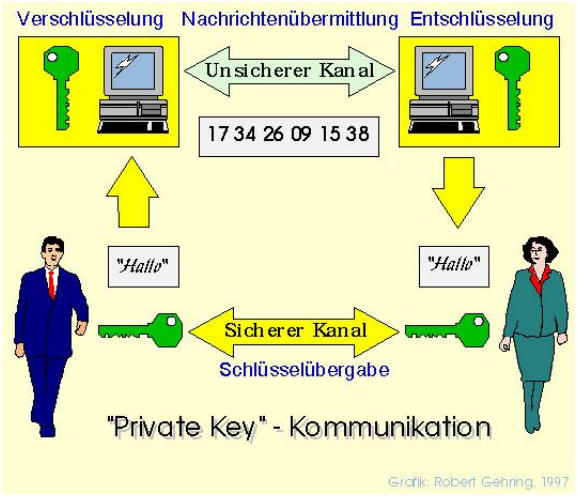

Abbildung 2.7: Eine symmetrische Verschlüsselung[12]

Die symmetrischen Verfahren verwenden normalerweise entweder die Datenblock- oder die Datenstrom-Verschlüsselung. Bei der Datenblock-Verschlüsselung werden jeweils komplette Blöcke einer bestimmten Größe mit einem Schlüssel verschlüsselt. Bei der Datenstrom-Verschlüsselung wird ein Datenstrom so lange fortlaufend verschlüsselt, bis der Eingangsdatenstrom vollständig verarbeitet wurde. [2]

- **Der Data Encryption Standard (DES)**

  DES wurde bereits Mitte der siebziger Jahre von IBM entwickelt. Er ist ein Standard des US-amerikanischen National Bureau of Standards (NBS) und ist in der Federal Information Processing Standard Publication (FIPS-Pub) 46-2 beschrieben. DES ist ein Blockverschlüsselungs-Verfahren und ist bis heute noch das am weitesten verbreitete Verschlüsselungs-Verfahren. Der Schlüssel hat eine Länge von 64-Bit, aber jedes achte Bit dient einer Paritätsprüfung und trägt nicht zur Verschlüsselung bei. Heutzutage ist 56-Bit Schlüssellänge unsicher und mit spezieller Hardware der NSA oder einem Brute-Force-Angriff ist DES innerhalb weniger Stunden knackbar. [2]

- **Triple-DES**

  DES galt jahrelang als ein sicheres Verfahren, sein einziger Nachteil ist die geringe Schlüssellänge. Deshalb schaltet man drei DES-Verschlüsselungen mit zwei oder drei Schlüsseln hintereinander. Der Klartext wird mit dem ersten Schlüssel verschlüsselt, mit dem zweiten entschlüsselt und wieder mit dem dritten verschlüsselt. Deshalb ist Triple-DES 3 fach langsamer als Standard-DES, dafür ist es aber sicherer. Man muss beachtet, daß wenn der erste Schlüssel und der zweite oder der zweite- und der dritte Schlüssel gleich sind, dann funktioniert 3DES wie Standard-DES. Aufgrund der Schlüssellänge gilt Triple-DES derzeit noch als sicheres Verfahren im Gegensatz zum einfachen DES. [13]

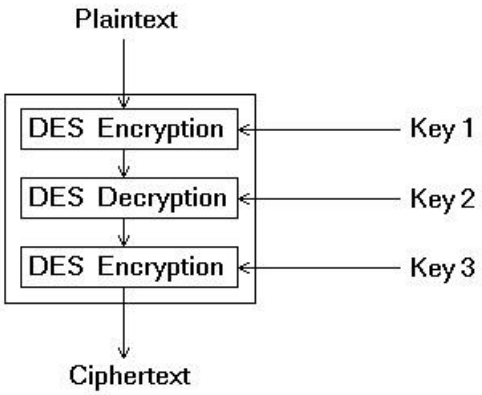

Abbildung 2.8: Triple-DES Verschlüsselung[12]

- **Advanced Encryption Standard (AES)**

  Das amerikanische National Institute of Standards and Technology(NIST) hat im Dezember 2000 den Rijndael-Algorithm als AES ausgewählt. Dieser Algorithmus wurde von Dr. Joan Daemen von Proton World International und von Dr. Vincent Rijmen von der katholischen Universität Leuven in Belgien

entwickelt. AES ist der Nachfolger für DES bzw. Triple-DES und ist ein Blockverschlüsselungsalgorithmus. [2] Rijndael ist ein Blockchiffrierverfahren. Die Blockgröße beträgt 128, 192 und 256 Bit, sowie variable Schlüssellängen mit 128, 192 und 256 Bit (AES-128, AES-192 bzw. AES-256). Nach Angaben des NIST würde eine Maschine, die einen DES-Schlüssel in einer Sekunde knackt, für einen 128-bit AES-Schlüssel 149 Billionen Jahre benötigen[14]

- **International Data Encryption Standard (IDEA)**

  IDEA wurde von Xueija Lai und James Massey entwickelt. IDEA ist ein Blockverschlüsselungsalgorithmus mit einer Blocklänge von 64 Bit und einer Schlüssellänge von 128 Bit. IDEA ist für nichtkommerzielle Anwendung lizenzfrei erhältlich. Bei kommerzieller Nutzung muss eine Lizenzgebühr an den Inhaber der Rechte(die Firma Ascom) gezahlt werden.[3] Wegen der großen Schlüssellänge kann IDEA heutzutage als sicherster Blockchiffrieralgorithmus angesehen werden.[15]

- **RC4**

  RC4 wurde im Jahr 1987 von Ron Rivest für die Firma RSA Data Security Inc.(RSADSI) entwickelt. Es handelt sich um eine Stromchiffrierung, die mit variabler Schlüssellänge arbeitet. Die variable Schlüssellänge kann bis zu 2.048 Bit lang sein. Jedes Zeichen wird einzeln verschlüsselt. Der Algorithmus war ganze sieben Jahre lang geheim, bis 1994 jemand anonym den Quellcode veröffentlicht hat. Dieser Algorithmus ist etwa fünf bis zehn mal so schnell wie DES.[3]

- **Blowfish**

  Blowfish ist eine Blockchiffrierung und wurde 1993 von Bruce Schneier entwickelt. Die Blockgröße beträgt 64-Bit, die Schlüssellänge ist variabel und kann bis zu 448 Bit reichen. Weiterhin ist Blowfish auf 32-Bit-Prozessoren wesentlich schneller als DES und bisher ist kein Makel am Blowfish-Algorithmus bekannt. [3]

2. **Asymmetrische Verschlüsselungsverfahren**

   Asymmetrische Verfahren verwenden zwei verschiedene Schlüssel. Auf der einen Seite ist der öffentliche Schlüssel (public key) zum Verschlüsseln(Chiffrierung) und auf der anderen Seite ist ein privater Schlüssel (private key) zum Entschlüsseln (Dechiffrierung).[2] Da ein Schlüssel öffentlich zugänglich gemacht werden muss, wird dieses Verfahren auch Public Key-Verfahren genannt. Basis der Asymmetrischen Verschlüsselungsverfahren sind Einweg-Funktionen. Dies sind Funktionen, die ohne größeren Aufwand berechenbar sind, bei denen aber die Berechnung der inversen Funktionen sehr schwierig ist. Wenn das Inverse einer Einweg-Funktion berechenbar sein soll, führt dies zu den Einweg-Funktionen mit Falltür(trapdoor onewayfunctions). Mit Hilfe einer geheimzuhaltenden Zusatzinformation ist die Berechnung des Inversen schneller machbar. Asymmetrische Verschlüsselungsverfahren sind etwa 1000 bis 10000 Mal langsamer als symmetrische Verfahren. Wegen des geringeren Datendurchsatzes bei der Ver- und Entschlüsselung werden sie nur zur Authentikation und zum Austausch des Schlüssels für die symmetrische Verschlüsselungsverfahren benutzt.[3] Die bekanntesten Public-Key-Verfahren sind RSA und Diffie-Hellman.

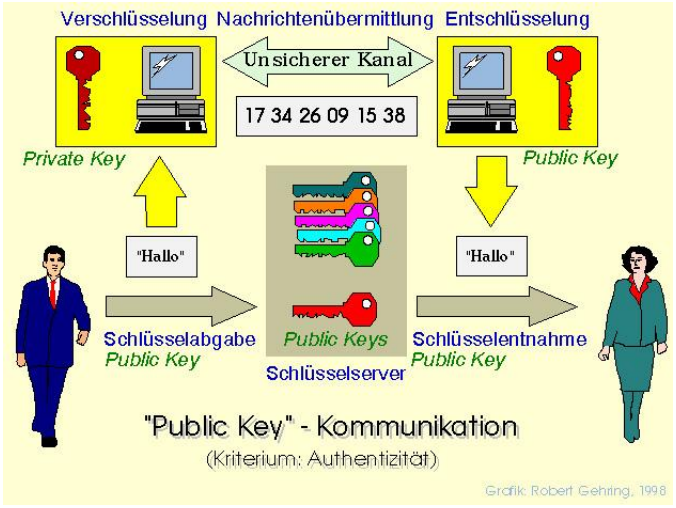

Abbildung 2.9: Asymmetrische Verschlüsselung[12]

- **Das Diffie-Hellman-Verfahren**

  Diffie-Hellman-Verfahren wurde 1976 von Whitfield Diffie und Martin Hellman entwickelt. Dieses sehr altes Verfahren kann nicht zum Ver- und Entschlüsseln von Daten verwendet werden. Es kann nur zur Erzeugung von symmetrischen Schlüsseln benutzt werden.[2] Das Verfahren beruht auf dem bis heute ungelösten mathematischen Problem des diskreten Logarithmus.[16]

- **Das RSA-Verfahren**

  Das RSA-Verfahren ist nach Entdeckern Ronald Rivest, Adi Shamir und Leonard Adleman vom Massachusetts Institute of Technology benannt. Der Unterschied zum Diffie-Hellman-Verfahren ist, daß man mit RSA Daten ver- und entschlüsseln kann. Die Funktion von RSA beruht auf dem mathematischen Problem, daß die Zerlegung einer großen Zahl in ihre Primfaktoren sehr aufwändig ist.[2]

3. **Hashfunktionen**

   Hashfunktionen sind keine Funktionen, zum Ver- oder Entschlüsseln. Sie berechnen einen kurzen Ausgangswert fester Größe(Hashwert) eines beliebig langen Eingangswerts. Dieser Hashwert kann als eine Art Prüfsumme bezeichnet werden. Der Hashwert wird digital unterschrieben und an die Nachricht angehängt. Der Empfänger wendet auf die Nachricht dieselbe Hashfunktion an und vergleicht das Ergebnis mit dem mitgeschickten durch die digitale Unterschrift des Senders authentifizierten Hashwert. Da eine kleine Änderung in der Datenmenge große Auswirkungen auf den Hashwert hat, kann man Manipulationen an den Originaldaten sofort merken. Damit kann die Integrität einer Nachricht sichergestellt werden. [18]

   Message Digest (MD5)

   MD4 (Message Digest No. 4), dokumentiert in RFC 1320, wurde von Ron Rivest (RSA Data Security) entwickelt. Dieser Algorithmus erzeugt einen 128-bit Hash-Wert in 3 Runden mit einer extrem schnellen Funktion. Leider führt diese Geschwindigkeit auch dazu, dass MD4-verschlüsselte Daten schnell wieder entschlüsselt

werden können, deshalb wurde ein Nachfolger von MD4 entwickelt. MD5 ist ein verbesserter Nachfolger von MD4 und er wurde ebenfalls von Ron Rivest entwickelt. MD5 erzeugt aus einem Eingangswert beliebiger Länge einen 128-Bit-Hashwert in 4 Runden. Dieses Verfahren findet man nicht nur in IPSec, sondern auch in vielen anderen Protokollen wie CHAP, L2TP.[3]

Secure Hash Algorithm (SHA-1)

Am Anfang der neunziger Jahre entwickelte der National Institute of Standards and technology(NIST) zusammen mit der National Security Agency(NSA) den Secure Hash Algorithm für den Einsatz mit dem Digital Signature Standard. Der Secure Hash Algorithm(SHA) erzeugt einen 160 Bit langen Hashwert und gilt als sehr sicher.[3]

Hash-based Message Authentication Code (HMAC)

Hash-based Message Authentication Code(HMAC) ist kein Hashalgorithmus, sondern ein Mechanismus zur Authentifizierung von Nachrichten, der kryptografische Hashfunktionen wie MD5 oder SHA verwendet. Die kryptografische Stärke des HMAC hängt von den Eigenschaften der zugrunde liegenden Hashfunktion ab.[17]

2.4 Zuordnung der VPN-Protokolle zu den 7 Schichten des ISO/OSI-Referenzmodelles

Heutzutage gibt es auf dem Internet-Markt viele Sicherheitsprotokolle, die zur Realisierung eines VPN verwendet werden können. Die Protokolle können in die verschiedenen Schichten des ISO/OSI-Referenzmodelles eingeordnet werden.

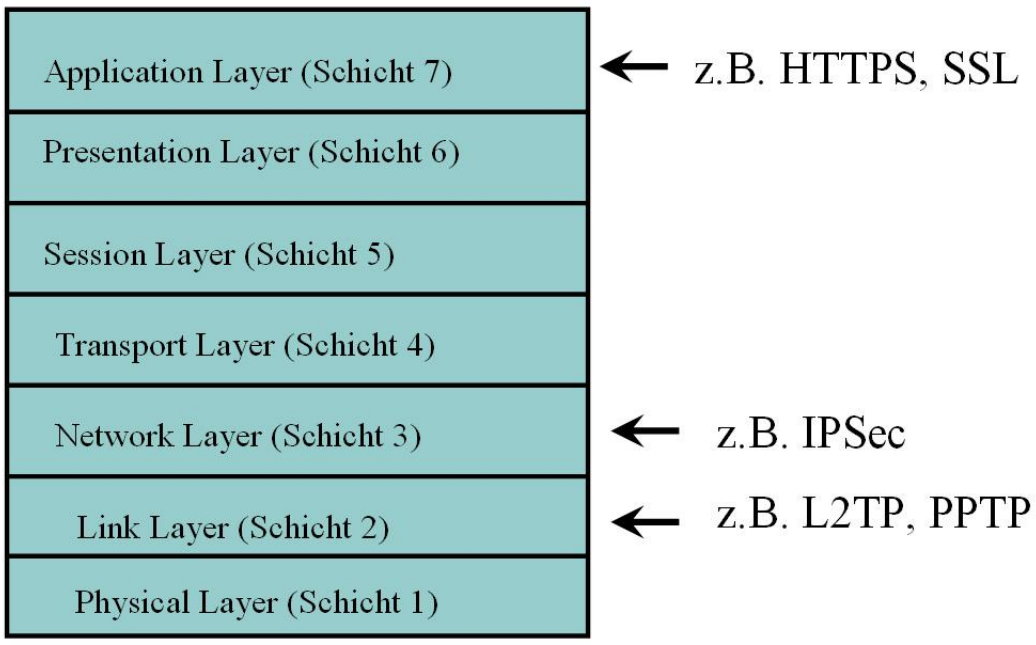

Abbildung 2.10: OSI-Schichten

Durch diese Zuordnung lassen sich drei unterschiedliche VPN-Typen klassifizieren. Es sind die VPN der Anwendungsebene (Application-Layer), der Netzwerkebene (Network-Layer) und der Netzwerkverbindungsebene (Link-Layer). In dieser Arbeit werden die häufig verwendeten Protokolle betrachtet: PPTP,L2TP,IPSec und SSL.

1. **PPTP : Point-to-Point-Tunneling-Protokoll**

   Das Punkt-zu-Punkt Tunneling Protokoll(PPTP) wurde von verschiedenen großen Firmen wie Microsoft, Ascend Communications, 3Com und anderen zusammen entwickelt. PPTP ist eine Erweiterung des Point-to-Point Protocol (PPP) und es ist von IETF(Internet Engineering Task Force) als Standardprotokoll für das Internet-Tunneling vorgeschlagen . Das PPTP wurde ursprünglich für "Remote Access Server" entwickelt und Microsoft hat PPTP erstmals in Windows NT4.0 integriert. PPTP kapselt PPP-Pakete in IP-Pakete, dadurch können andere Pakete wie IPX- oder NetBUI-Paketen getunnelt werden. Aber je Kommunikationspaar kann nur ein Tunnel aufgebaut werden. Der Tunnel besteht aus zwei Übertragungskanälen: einer für die Nutzdaten, der andere für die Kontrolle der Verbindung. Über diesen Kontroll-Kanal signalisieren sich die Systeme Verbindungsauf- und -abbau. Für die Authentifizierung der Kommunikationspartner setzt PPTP auf die von PPP bereitgestellten Mechanismen. Diese sind das Password Authentication Protocol (PAP) oder das Challenge Handshake Protocol (CHAP). Die Verschlüsselung der Daten erfolgt über RC4-Verfahren entweder mit einem 40-Bit-Schlüssel für den nicht amerikanischen Markt, oder 128-Bit-Schlüssel. USA-Exportgesetze stufen 128-Bit-Verschlüsselungstechnik als militärische Technologie ein.[20]. Obwohl man PPTP noch heute verwendet, findet keine Weiterentwicklung mehr statt. Wegen der nicht genügenden Sicherheit bei der Authentifizierung und Datenverschlüsselung wird PPTP immer weniger benutzt und durch andere Verfahren eingesetzt.

2. **L2TP : Layer-2-Tunneling-Protokoll**

   Layer-2-Tunneling-Protokoll wurde von IETF als ein Standardprotokoll veröffentlicht. Es arbeitet auf der Schicht 2 des ISO/OSI-Schichtenmodells. L2TP ist eine Verbesserung und Weiterentwicklung zweier nicht standardisierter Protokolle, des Point-to-Point-Tunneling-Protocol (PPTP) und des Layer-2-Forwarding (L2F). L2TP ist ein reines Tunneling-Protokoll auf PPP-Ebene, das auch andere Protokolle als nur IP einkapseln kann. In L2TP sind keine Sicherheitsfunktionen wie Authentifizierung oder Datenverschlüsselung verfügbar. Innerhalb des Tunnels überträgt L2TP zwei unterschiedliche Datenströme. Zum einen die Kontrollnachrichten, zum anderen die Nutzdaten. Während der Kontrollkanal über eine gesicherte Verbindung gesendet wird, stellt L2TP keine Möglichkeit zur Verfügung, die Nutzdaten zu sichern. Sicherheitsmaßnahmen werden auf anderen Ebenen eingesetzt wie auf der Netzwerkebene(z.B. IPSec) oder auf applikationsnahen Ebenen(z.B. SSL).[2] L2TP erlaubt eine Authentifizierung auf der Basis von PAP oder CHAP und es bietet die Möglichkeit, multiple Tunnels aufzubauen. Als Bestandteil von Windows 2000 hat die Benutzung des L2TP weiter zugenommen.

3. **IPSec**

   IPsec war ursprünglich für IP Version 6 geplant. Für IPv6 bietet IPSec die Verschlüsselung auf Netzwerkebene, also auf der Schicht 3 des OSI-Schichtenmodells. IPSec ist eine neue Technik, die PPTP ablösen soll, weil sie mehr Sicherheitsmaßnahmen als PPTP anbieten. Im Gegensatz zur IP-Version 4 werden zwei zusätzliche Header gegeben, die zu Zwecken der Authentifizierung des Absenders und Verschlüsselung definiert werden. Die zusätzlichen Header bei IPv6 sind Authentication Header (AH) und Encapsulated Security Payload (ESP). IPSec wurde aber auch mit IP-Version 4 implementiert, wobei eine Erweiterung des normalen IP-Header stattfindet. Die Authentifizierung des Absenders wird durch RSA, Pre-Shared Keys oder X-509 Zertifikaten sichergestellt. Dazu werden Schlüssel der Kommunikationspartner per Internet Key Exchange(IKE) ausgetauscht. Die in IPSec häufig verwendeten Algorithmen zur Verschlüsselung der Daten sind: DES mit einem 56 Bit langen Schlüssel, Triple-DES mit 168 Bit-Schlüssel, IDEA mit einem 128 Bit langen Schlüssel oder Blowfish mit 448 Bit großem Schlüssel.

   IPSec bietet dem Benutzer zwei Betriebsmodi, Der Tunnel Modus und der Transportmodus. Diese können für unterschiedliche Einsatz-Szenarien eingesetzt werden.

   (a) Tunnelmodus

   Der Tunnelmodus wird benutzt, um das gesamte IP-Paket in die Nutzdaten eines IPSec-Pakets einzupacken. Ein IPSec-Header wird vor das IP-Paket und zusätzlich ein weiterer IP-Header eingefügt. Der Tunnelmodus kann in allen drei Einsatzszenarien verwendet werden

   (b) Transportmodus

   Im Transportmodus wird der IP-Header des ungesicherten IP-Pakets übernommen und nur sein Datenteil verändert. Der IPSec-Header wird zwischen dem IP-Header und dem Header des übergeordneten Protokolls eingefügt. Der Transportmodus kann ausschließlich in Host-to-Host-Einsatzszenario benutzt werden, da die Kommunikationsendpunkte auch gleichzeitig die Sicherheitsendpunkte sind.[2]

   - **Das Authentication-Header-Protokoll (AH)**

     Authentication Header ist ein IPSec-Protokoll, das die Datenvertraulichkeit anbietet. Die Aufgabe des AH sind die Authentifizierung der Datenpakete und die Datenintegrität. Ausserdem bietet es auch den Schutz vor Replay-Angriffen. Zum Schutz der Datenintegrität und zur Authentifizierung werden Hash-Algorithmus: SHA oder MD5 eingesetzt. Das AH wird im Tunnelmodus und im Transportmodus eingesetzt wie in Abb.2.11 gezeigt.

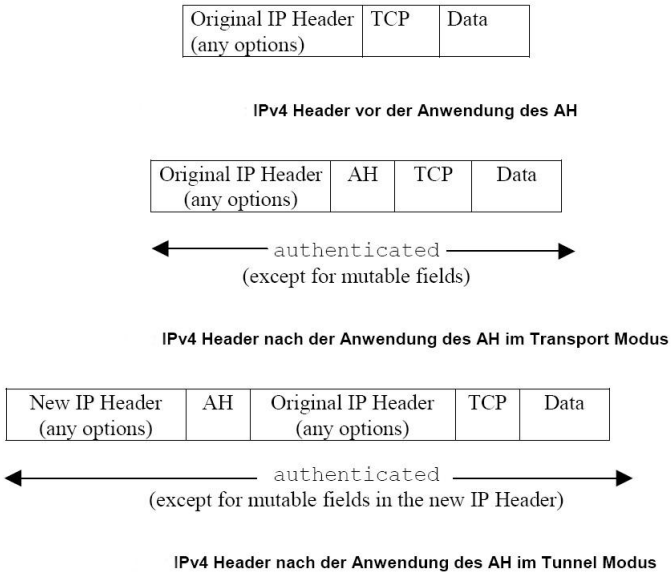

Abbildung 2.11: Das Authentication-Header-Protokoll(AH)[28]

- **Das Encapsulating-Security-Payload-Protokoll (ESP)**
  Das Encapsulating-Security-Payload-Protokoll bietet fast gleiche Sicherheitsfunktionen wie beim AH, aber es bietet zusätzlich noch die Datenverschlüsselung. Dazu wird ein Teil des IP-Pakets verschlüsselt. Im Tunnelmodus werden Applikations-Header, TCP/UDP-Header und Daten verschlüsselt, während im Transportmodus nur TCP/UDP-Header und Daten verschlüsselt werden. Wie beim AH benutzt ESP zum Schutz der Datenintegrität und zur Authentifizierung einen Hash-Algorithmus, SHA oder MD5.

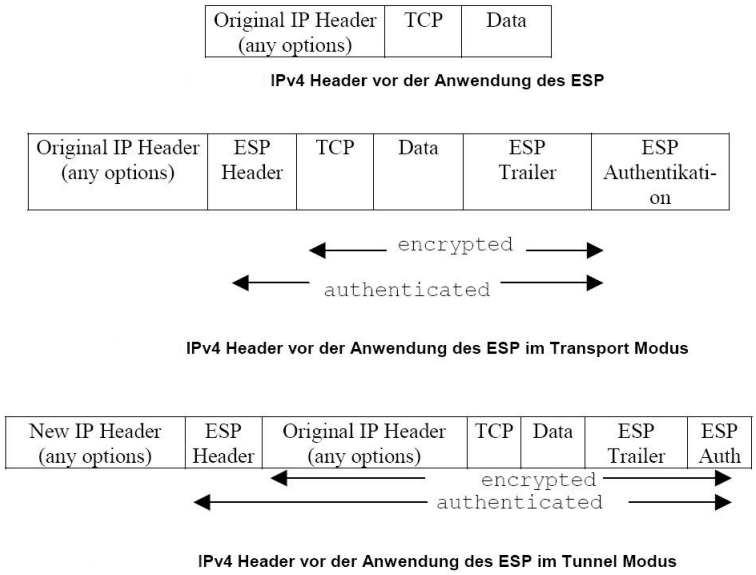

Abbildung 2.12: Das Encapsulating-Security-Payload-Protokoll(ESP)[28]

	IPSec	L2TP	PPTP
Protokolltyp	**Layer3**	**Layer2**	**Layer2**
IPv6 kompatibel	**ja**	**ja**	**ja**
Authentifizierung Verbindungsaufbau	**Preshared Keys,**	**PAP, CHAP, EAP RSA, X.509**	**PAP, CHAP**
Verschlüsselung	**3DES, AES**	**-**	**RC4, DES**
Max. Schlüssellänge	**168 Bit (3DES) 256 Bit (AES)**	**je nach Applikation**	**128 Bit (RC4) 56Bit (DES)**
Paket-Authentifizierung	**ja**	**nein**	**nein**
Prüfung der Datenintegrität	**ja**	**nein**	**nein**

Abbildung 2.13: Übersicht[1]

4. **SSL-VPN**

   Das SSL-Protokoll wurde ursprünglich von Netscape entwickelt. Die derzeit aktuellste Version des SSL-Protokolls ist die Version 3.0. Zwischen zwei Rechnern können mit SSL-Protokoll ein oder mehrere Sitzungen auf der aktuellen Basis einer 128bit Verschlüsselung aufgebaut werden. Das Protokoll läuft im OSI-Schichtenmodell zwischen der Anwendungs- und Transportschicht und bietet Authentizität, Vertraulichkeit und Unversehrtheit der Daten wie IPSec. SSL benutzt sowohl asymmetrische Verschlüsselungsverfahren als auch symmetrischen Verschlüsselungsverfahren. Das SSL Protokoll besteht aus mehreren Protokolle[23](siehe Abb. 2.14) Das Handshake Protokoll dient dazu, einen Verschlüsselungsalgorithmus auszuwählen und die Authentifizierung eines Clients mit einem Server zu ermöglichen. Die Authentifizierung des Kommunikationsteilnehmers wird mit Hilfe von Zertifikaten von Zertifizierungsinstanzen (Certification Authorities, CAs) realisiert. Die Daten höherer Protokolle werden durch SSL Record Protokoll fragmentiert, komprimiert und verschlüsselt und an das untere Transportprotokoll weiter geleitet. Dieses Record Protokoll realisiert die Vertraulichkeit, Integrität und Authentizität der zwischen Client und Server versendeten Daten. Das weitere Protokoll, Change Cipher Spec-Protokoll, regelt den Wechsel zu einem anderen Verschlüsselungsalgorithmus. Das Alert Protokoll dient dazu, Fehlermeldungen weiterzuleiten und das Application Data Protokoll reicht die Daten des oberen Protokolls an das Record Protokoll weiter. Der Vorteil des SSL-Protokolls ist die Möglichkeit, jedes höhere Protokoll auf Basis des SSL Protokolls zu implementieren. Damit ist eine Unabhängigkeit von Applikationen und Systemen gewährleistet. SSL-VPN wurde von der Industrie, als neue Lösung für Remote Access bzw. Szenario 3(Gateway-to-Host), entwickelt.

   Für den Schlüsselaustausch während des Handshakes stehen die Verfahren RSA und Diffie-Hellman zur Verfügung. Zur Verschlüsselung der Daten können folgende Algorithmen verwendet werden: RC4, DES und IDEA.

   Vorteil bei dieser Lösung ist der uneingeschränkten Fernzugriff auf Unternehmens-

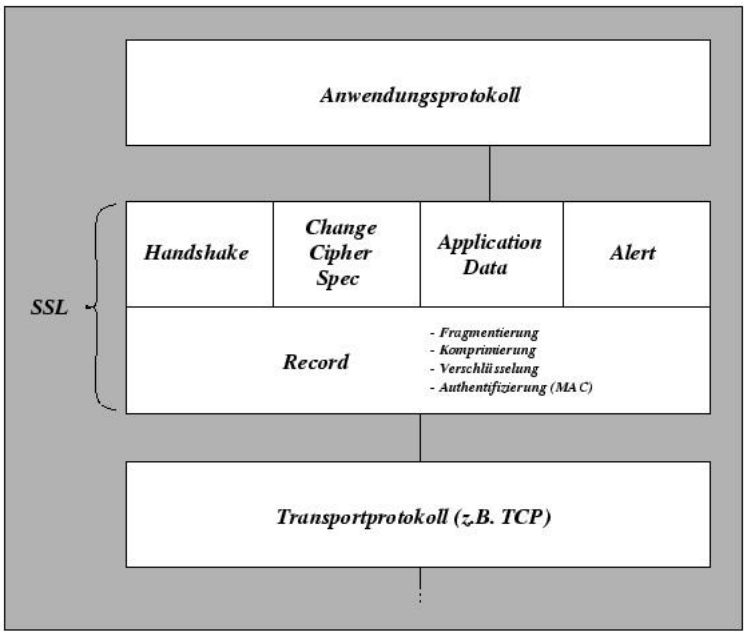

Abbildung 2.14: Das SSL-Schichtenmodell[23]

applikationen über das Internet von einem Endgerät. Bei diesem Vorteil besteht aber dennoch eine Gefahr, wenn das Endgerät nicht sicher ist z.B. ein Computer im Internetcafe oder ein von Viren verseuchter Computer.

SSL VPNs können in 3 Kategorien eingeordnet werden [24]:

- Application layer proxy
- Protocol redirectors
- Remote control enhancers

Application layer proxy

Application layer proxy ist die einfachste Form des SSL-VPN, weil sie eine Verschlüsselungsschicht zwischen die Applikation(Schicht 7) und die darunterliegende Schichten des ISO/OSI Stacks legen. Dadurch werden alle Daten, die die Applikation sendet, verschlüsselt und alle, die sie empfängt, entschlüsselt. Von einem "Applikation Layer Proxy" spricht man, wenn der Secure Socket Layer (SSL) in die Applikation, z.B. einen Browser oder eine E-Mail Client integriert ist.Es wird häufig bei E-Banking und E-Commerce benutzt. Da man kein zusätzliche Programme für VPN auf dem Rechner installieren müssen, wird dieses Verfahren auch Clientless genannt.

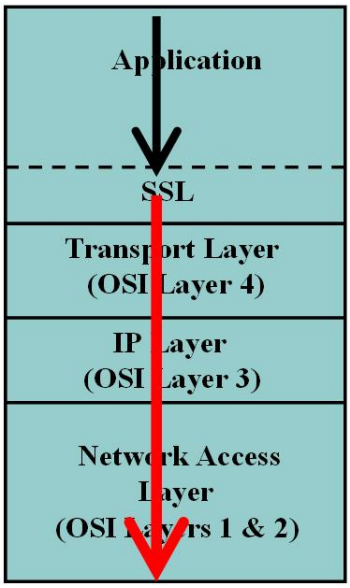

Abbildung 2.15: SSL-VPN : Application layer proxy

Protocol redirectors

Dieses Verfahren, Protocol redirectors, wird benutzt, wenn das SSL Protokoll nicht in die Applikation integriert ist. Bei diesem Verfahren muss ein SSL-Client Programm wie z.B. Java-Applets und Active-X-Controls heruntergeladen und installiert werden. Die Daten werden durch Port Forwarding von Schicht 3 zum SSL-Client umgeleitet und sie werden von dort weiter zu einem anderen Rechner gesendet(siehe Abb.2.16).

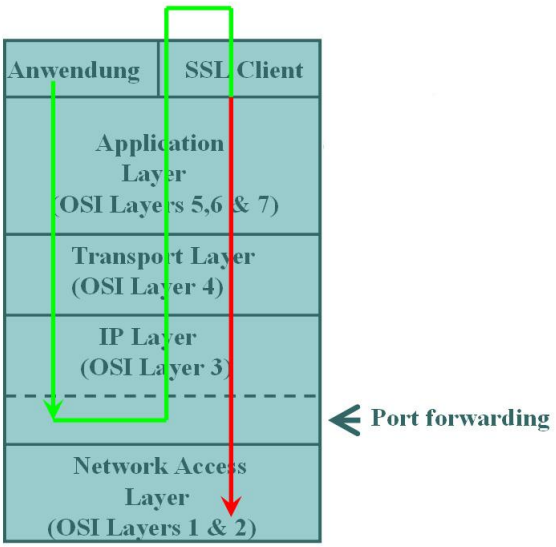

Abbildung 2.16: SSL-VPN : Protocol redirectors

Remote control enhancers

Remote control enhancers bietet SSL-Verbindung für existierende remote control Produkte, wie Windows Terminal Services. Die SSL-Funktionalität und Web-Browser

Unterstützung werden ins System integriert, dadurch kann jede Applikation, die auf einem solchen System läuft, die SSL-Funktion benutzen.

5. Implementierung der VPN-Protokolle

   Heutzutage werden verschiedene VPN-Lösungen von unterschiedlichen Herstellern auf den Markt gebracht. Sie werden auf verschiedenen Betriebssystemen implementiert wie Windows oder Linux. Eine bekannte Implementierung ist FreeS/WAN in Linux. Es ist eine Implementierung des IPSec-Protokolls. Eine andere ist OpenVPN. Das OpenVPN verwendet ein Standardprotokoll: Transport Layer Security (TLS)[27] und ist unter Linux implementiert. TLS ist der Nachfolger des SSL-Protokolls von Netscape und benutzt X-509 Zertifikaten zur Authentifizierung wie SSL. Die CIPE(Crypto IP Encapsulation) ist eine Implementierung, die sowohl unter Linux als auch unter Windows verfügbar ist. Die CIPE verpackt IP-Pakete verschlüsselt in UDP-Paketen und die Pakete werden über UDP-Ports weitergesendet. Zur Authentifizierung wird ein Public-Key-Verfahren eingesetzt.[27]

   Neben Softwareimplementierung, existieren heute auch mehrere VPN-Hardware-Lösungen. Eine bekannte Lösung ist das System der Firma Cisco. Cisco bietet verschiedene Hardwarelösungen mit integrierten VPN-Servern an.[1]. Wegen dem extra eingebaute Chip, der die Verschlüsselung der Daten übernimmt, bietet das System sehr hohe Performance. Ausserdem ist es recht resistent gegen Angriffe, da es ein Cisco-Betriebssystem benutzt. Allerdings muss ein spezielles Programm auf jedem Rechner im Netzwerk installiert werden. Eine andere Hardware-Lösung ist auf Netzwerkkarte z.B. Intel Pro/100S implementiert.[25]. Sie bietet keine komplette Lösung wie bei den VPN-Geräten von Cisco, sondern sie hilft der CPU bei der Datenverschlüsselung und sie wird für IPSec-VPN gebaut.

2.5 Zusammenfassung

Heute existieren viele verschiedene VPN-Lösungen für das Internet. Sie werden für unterschiedliche Einsatzbereiche entwickelt. Zur Zeit im Bereich des Mobile-Systems sind IPSec-VPN und SSL-VPN die neuste Technik. Obwohl SSL-VPN von der Industrie als die bessere Technik bezeichnet wird, kann es trotzdem IPSec-VPN nicht ersetzen. Beide Verfahren bieten gleichwertige Sicherheit, aber man braucht bei IPSec-VPN eine aufwendigere Konfiguration. Transparente Tunnel werden mittels IPSec-VPN aufgebaut. Da IP-Sec als Schicht 3 Verschlüsselung tiefer in das System eingreift als SSL-VPN auf Schicht 7, muß man dafür Administratorrechte besitzen und die Konfiguration ist aufwendiger. Bei SSL-VPN muß praktisch nur eine Applikation mit SSL-Unterstützung wie z.B. ein Web-Browser installiert werden. Kann allerdings eine vorhandene Applikation, da die Quelltexte nicht vorhanden sind, nicht um SSL oder TLS erweitert werden, so muß auf das Protocol Redirector Verfahren zurückgegriffen werden. Deshalb kann man heute nicht sagen, welche Lösung am besten oder für alle Fälle geeignet ist.

Literaturverzeichnis

[1] “A VPN-based Security Solution for WiFi Networks“, Studienarbeit von Ronny Nantke, Institut für Informationstechnische Systeme, Universität der Bundeswehr, Neubiberg, August 2003

[2] “VPN - Virtuelle Private Netzwerke“, Manfred Lipp, Addison-Wesley Verlag, 2001

[3] “Virtual Private Networks“, Dr. Markus a Campo Dr. Norbert Pohlmann, mitp-Verlag/Bonn, 2.Auflage, 2003

[4] “VPN in der Praxis“, Linux Magazin 10, 2003

[5] http://aib.de/nsa.htm

[6] http://vpnwww.uni-jena.de/vpnvortrag.html, ’Virtual Private Networks’, Vortrag an der Friedrich-Schiller-Universität Jena

[7] “Virtual Private Network : Mit sicherem Tunnel durchs Internet FreeS/WAN “, Diplomarbeit von Olivier Gärtner und Berkant Uenal, November 1999

[8] http://www.webagency.de/infopool/internetwissen/internetabc_pq.htm

[9] http://www.webagency.de/infopool/internetwissen/internetabc_c.htm

[10] http://www.teco.edu/ zimmer/vpn/node11.html, RADIUS-Remote Authentication Dial-In User Service, Tobias Zimmer

[11] http://www.bintec.de/de/solution/security/vpn/pdf/DS/sw/Netzsicherheit_de.pdf

[12] http://ig.cs.tu-berlin.de/ap/rg/1998-04/grundlagen.html, “Grundlagen : digitale Signaturen“, Diplomarbeit von Robert Gehring

[13] http://www.tropsoft.com/strongenc/des3.htm, Triple DES Encryption

[14] http://www.nist.gov/public_affairs/releases/aesq&a.htm, Advanced Encryption Standard (AES), Questions and Answers

[15] http://ig.cs.tu-berlin.de/ap/rg/1998-04/glossar/i-terms/idea.html, IDEA -International Data Encryption Standard

[16] http://www.t-lan.de/GLOSSAR/begriffe/diffie-hellman-verfahren.htm, Diffie-Hellman-Verfahren

[17] http://www.telekom.de/dtag/tklex/tklex_cda_index/1,13403,5409,00.html

[18] http://www.networks.siemens.de/solutionprovider/_online_lexikon/8/f004718.htm

[19] http://www.kom.id.ethz.ch/datkom/vpn/was_ist_vpn.html

[20] http://www.lanline.de/html/lanline/lexikon/lex/pptp.htm, PPTP

[21] http://home.t-online.de/home/TschiTschi/ssl.htm#tunnel, SSL

[22] http://developer.netscape.com/docs/manuals/security/sslin/contents.htm#1041640, Introduction to SSL

[23] http://netsplit.de/crimson/node105.html, Das Schichtenmodell

[24] http://networknewz.com/networknewz-10-20031201SSLVPNinDetail.html, SSL VPN in detail

[25] http://www.intel.com/network/connectivity/products/pro100s_srvr_adapter.htm, Intel Pro/100S

[26] http://www.funkschau.de/heftarchiv/pdf/2003/fs2303/fs0323026.pdf, Eine sichere Verbindung

[27] Linux Magazin 10/2003: “VPN in der Praxis“

[28] http://www.informatik.uni-hamburg.de/RZ/lehre/18.415/seminararbeit/10_VPN.pdf

Kapitel 3

Realtime Networking in Wireless and Wired Networks

Ronny Nantke

Diese Arbeit beschäftigt sich mit Realtime Networking in Wireless and Wired Networks. Dabei sollte untersucht werden, in wie weit sich diese beiden Netzwerktechniken für Soft- und Hardrealtime Anforderungen eignen. Berücksichtigt wurden insbesondere die Schicht 2 des OSI/ISO Referenzmodells, da das Mediumzugriffsverfahren die Eigenschaften im Hinblick auf Reaktionszeit, sowie Garantien für die Zuteilung des Mediums bestimmen. Ein weiteren Schwerpunkt stellen die Realtime Protokolle der Schicht 4 da. Dort wurde insbesondere die Tauglichkeit für Wireless LANs betrachtet.

Inhaltsverzeichnis

3.1 Einleitung 55
3.1.1 Unterschied zwischen Wireless und Wired Networks 55
3.1.2 Definition Soft- und Hardrealtime 56
3.1.3 Das idealisierte Echtzeitnetzwerk 57
3.1.4 Das OSI/ISO Referenzmodell 58
3.2 Die Netzwerktechniken im Vergleich (Schicht 2) 58
3.2.1 Allgemeine Multiplexverfahren 58
3.2.2 Vergleich der MAC Frames 59
3.2.3 MEDIA ACCESS CONTROL Protokolle in Ethernet Kabelnetzwerken 60
3.2.4 MEDIA ACCESS CONTROL Protokolle in Funknetzwerken . 63
3.2.5 Zusammenfassung der Mediumzugriffsverfahren 67
3.3 Realtimeprotokolle 68
3.3.1 Realtime Protocol (RTP nach IETF RFC3550) 68
3.3.2 Realtime Transport Control Protocol (RTCP nach IETF RFC3550) 69
3.3.3 RTnet der Universität Hannover 69
3.3.4 Protokolle Wireless tauglich? 70
3.3.5 Kommerzielle Realtime Protokolle 71
3.4 Existierende Systeme 71
3.4.1 Roboter im Kaufhaus (RoBoKa) 71
3.5 Zusammenfassung 72
3.5.1 Fazit 72
3.5.2 Ausblick 73

3.1 Einleitung

Die moderne Kommunikationsindustrie setzt aus Kosten- und Leistungsgründen für die Zukunft verstärkt auf Video- und Voice over IP Lösungen [1]. Dabei kommen im Home- und Officebereich kabellose und Kabelnetzwerke zum Einsatz. Allerdings möchte der Kunde, wie beim herkömmlichen Telefonieren, einen Qualitäts- und Verfügbarkeitsstandard. Darüber hinaus gibt es aber auch Industrieanwendungen: So werden Fertigungsanlagen und Roboter über Realtimeapplikationen gesteuert. Benötigen Steuerungsbefehle zu lange Zeit, kann es zu Fehlproduktionen oder gar teuren Beschädigungen kommen.

Aber auch in weitaus kritischeren Bereichen werden Echtzeitanwendungen eingesetzt: So muss die rechtzeitige Abschaltung eines Kernkraftwerks gewährleistet werden. Bei der Flugbahnberechnung einer Trägerrakete wird die Absprengung der Zusatztanks ebenfalls exakt geplant. Kommt es bei der Übertragung der entsprechenden Befehle vom Zentralcomputer zu den jeweiligen Steuereinheiten zu Verzögerungen, bedeutet das den Verlust der Rakete samt Ladung, was einen wirtschaftlichen Schaden von mehreren hundert Millionen Euro zur Folge hat [2].

Aber nicht nur im zivilen Umfeld spielen Realtime Netzwerke eine Rolle: Aufklärungsdrohen erhalten per Funksignalen ihre Steuerungsbefehle. Gleichzietig senden sie Video-, Infrarot- und GPS Informationen an die zugehörige Bodenstation. Von dort aus werden sie in Echtzeit an Kampfeinheiten im Einsatzgebiet weitergeleitet. Erreichen dabei Informationen zu spät, oder gar fehlerhaft den Empfänger, gehen die Konsequenzen von Eigenverlusten bis zu hohen Kolatteralschäden [3].

3.1.1 Unterschied zwischen Wireless und Wired Networks

Um die beiden Netzwerktechniken untersuchen zu können, betrachtet man erstmal die Unterschiede bzw. die Vor- und Nachteile. Während Kabelnetzwerke in der Regel Kupfer- oder Glasfaserkabel als Übertragungsmedium nutzen, werden im Wireless LAN die Bits mittels elektromagnetischer Wellen über den Äther verschickt. Das bewerkstelligt auch den großen Vorteil der WLANs, dass die User innerhalb eines Hotspots völlige Bewegungsfreiheit haben. Kabelnetzwerke verfügen nur über den beschränkten Raum, den die Kabellänge zulässt, wobei hier auch noch darauf geachtet werden muss, dass man Kabel nicht einfach quer durch den Raum legen kann.

Das führt gleich dazu, dass man Unterschiede in der Topologie feststellen kann: Während die Anordnung der Mitglieder eines WLANs völlig beliebig ist, sind Kabelnetze an eine Baum-, Ring- oder Linienstruktur gebunden.

Inwieweit sich Netzwerktechniken durchsetzen beruht insbesondere auf dem Kostenfaktor: WLAN Adapterkarten kosten um die 40,- Euro und die benötigten Accesspoints um die 100,- Euro. Kupfergebundene Lösungen liegen mit 10,- Euro für die Netzwerkkarte und 30,- Euro für einen Hub da deutlich günstiger.

Zumal sie mit bis zu 1000 Mbit im Gegensatz zu 54 Mbit von Wireless LANs eine deutlich größere Bandbreite bieten. Allerdings kann das Kabelverlegen auf größeren Strecken oder wegen baulichen Maßnahmen den Kostenpunkt von Kabelnetzwerken in die Höhe treiben.

Ein weiterer Schwachpunkt von WLANs ist der hohe Aufwand, die Übertragungen abzusichern. Bei Kabelnetzwerken muss eine direkte Verbindung bestehen, um den Datenverkehr abzuhören oder zu beeinflussen. Bei WLANs reicht es jedoch, sich innerhalb des Sendebereichs eines Hotspots mit einer entsprechenden Adapterkarte aufzuhalten.

WLANs bieten durch ihre Flexibelität einen Vorteil: Ist die Hardware richtig konfiguriert, so kann man das Netzwerk abbauen, und ohne größeren Aufwand an einem anderen Ort wieder in Betrieb nehmen, vorausgesetzt, dort gibt es keine Störquellen.

3.1.2 Definition Soft- und Hardrealtime

Bevor man die Echtzeiteigenschaften der beiden Netzwerktechniken untersucht, muss erstmal der Begriff Realtime geklärt werden: Generell gilt, es wird ein Zeitpunkt t_{min} und ein Zeitpunkt t_{max} festgelegt. Innerhalb dieses Zeitraumes sollten die Daten eintreffen. Dabei unterscheidet man aber zwei Kategorien: weiche (Soft) und harte (Hard) Echtzeitbedingungen. Während für weiche Echtzeitbedingungen die Grenzen t_{min} und t_{max} eher die optimale Zeitspanne bestimmen, müssen bei harten Echtzeitappliaktionen die Grenzen streng eingehalten werden, und dürfen weder unter- noch überschritten werden. Das Ganze lässt sich durch folgende Grafik veranschaulichen 3.1:

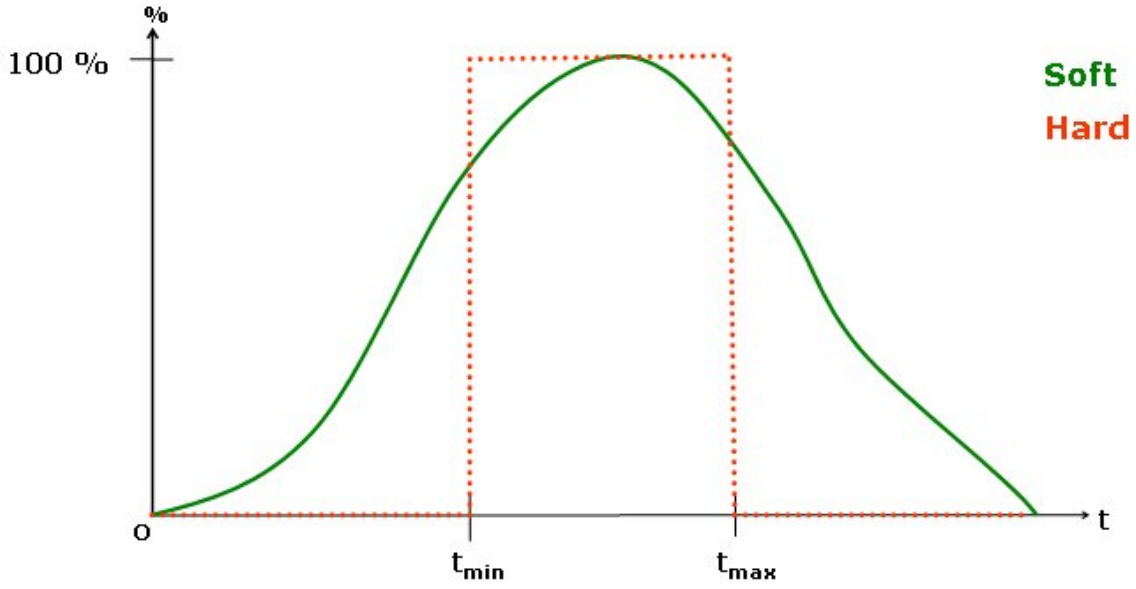

Abbildung 3.1: Soft- und Hardrealtime im Vergleich

Ein nicht Einhalten der Zeitspanne führt bei Softrealtime in der Regel zur Minderung der Übertragungsqualität. Bei Hardrealtime hingegen führt ein Verletzen der Zeitpunkte zur Zerstörung des Systems und wird somit nicht toleriert.

Anwendungsbeispiele für Softrealtime sind Bild- und Sprachübertragung, Videostreaming, Computerspiele, usw. Um die Übertragungsqualität zu erhöhen, verwenden die Applikationen kleine Zwischenspeicher. Dadurch können kleinere Unregelmäßigkeiten aufgefangen werden. Beispiele für Hardrealtime Systeme sind Anwendungen in der Raumfahrt, Steuerung von Kraftwerken und Industrieanlagen.

3.1.3 Das idealisierte Echtzeitnetzwerk

Nachdem der Begriff Realtime geklärt wurde, kann man ein idealisiertes Netzwerk für Realtimeumgebungen aufstellen:

Reaktionszeit: Möchte eine Station ein Paket senden, sollte das Medium sofort zur Verfügung stehen. Dies ist aber in einem herkömmlichen Netzwerk nicht immer möglich. Daher sollte zumindest eine maximale Zeit feststehen, in der auf das Medium zugegriffen werden kann.

Bandbreite: Gute Video- und Audiosignale benötigen immer mehr Bandbreite. Daher wäre eine unbegrenzte Bandbreite wünschenswert, aber nicht realisierbar. Daher sollte zumindest genügend Banbreite zur Verfügung stehen, um alle Teilnehmer mit der gewünschten Qualität versorgen zu können.

Garantien: Für harte Echtzeitapplikation gelten sehr strenge Zeitintervalle. Das idealisierte Netzwerk kann diese Zeitpunkte natürlich nicht exakt erreichen. Daher werden zumindest Garantien über kleine Zeitfenster gefordert und insbesondere Zeitpunktobergrenzen, die nie überschritten werden.

Signalausbreitung: Beginnt eine Station zu senden, sollte der Empfänger ohne Zeitverzug die Informationen empfangen. Dies ist aber gerade bei größeren Entfernungen nicht möglich, da sich die Signale maximal mit Lichtgeschwindigkeit c ausbreiten. Des weiteren werden gerade in Wide Area Networks (WANS) Komponenten wie Router und Switches benötigt, die die Pakete kurz zwischenspeichern. Zusätzlich bestehen bei WANS mehrere Möglichkeiten ein Paket von Station A zur Station B zu transportieren. Daher sollte immer der schnellste Weg zwischen zwei Punkten gewählt werden.

MAC Frames: Im idealen Netzwerk sollten nahezu nur Nutzdaten übertragen werden. Da aber für die einzelnen Schichten des ISO/OSI Referenzmodells zusätzliche Informationen benötigt werden, müssen die zusätzlichen Informationen der MAC Frames möglichst gering ausfallen.

Datengültigkeit: Gerade für harte Realzeitsysteme ist eine Vorraussetzung, dass die empfangenen Informationen korrekt sind. Es darf zwischen den Stationen nicht zu Übertragungsfehlern kommen. Falls doch muss zumindest gewährleistet sein, dass Fehler erkannt und eventuell sogar beseitigt werden können.

3.1.4 Das OSI/ISO Referenzmodell

Diese Arbeit untersucht beide Netzwerktechniken insbesondere im Hinblick auf die Schicht 2, MAC Layer, und Schicht 4, Transport Layer. Die Bitübertragungsschicht hat im Hinblick auf Realtime Eigenschaften keine große Bedeutung. Die Einordnung kann man in der Grafik 3.2 sehen.

Schicht 7	**Application Layer / Anwendungschicht**
Schicht 6	**Presentation Layer / Präsentationschicht**
Schicht 5	**Session Layer / Sitzungschicht**
Schicht 4	**Transport Layer / Transportschicht**
Schicht 3	**Network Layer / Vermittlungsschicht**
Schicht 2	**Data Link Layer / Sicherungsschicht**
Schicht 1	**Physical Layer / Bitübertragungsschicht**

Abbildung 3.2: Einordnung der zu untersuchenden Schichten

3.2 Die Netzwerktechniken im Vergleich (Schicht 2)

3.2.1 Allgemeine Multiplexverfahren

Möchten mehrere Stationen über dasselbe Medium Informationen übertragen, müssen Verfahren zur Verteilung des Mediums benutzt werden. In diesem Abschnitt werden die sogenannten Multiplexverfahren untersucht, insbesondere auf ihre Vor- und Nachteile für Echtzeitumgebungen. Dabei gilt als Anmerkung, dass die einzelnen Verfahren beliebig miteinander kombinierbar sind.

Space Division Multiplex (SDM)

Dabei wird das Medium durch eine räumliche Unterteilung aufgeteilt, so dass sich die Stationen nicht gegenseitig stören. Beim Einsatz von normalen Antennen bietet SDM die volle Bandbreite, allerdings nur für zwei Teilnehmer. Optimieren kann man die Teilnehmeranzahl durch den Einsatz von Richtantennen. Dadurch wird allerdings die Mobilität der Stationen stark eingeschränkt. Außerdem steigen dadurch die Kosten, da für jede benötigte Verbindung zwei Antennen vorhanden sein müssen [7].

Time Division Multiplex (TDM)

Bei TDM wird das Medium nacheinander den einzelnen Stationen zugeteilt. Dies kann durch eine feste Vergabestrategie erfolgen, was aber einen Synchronisationsmechanismus voraus setzt. Oder es wird ein dynamisches Vergabesystem verwendet, bei dem nur Stationen, die auch senden wollen, sich um das Medium bemühen. TDM bietet zwar die Möglichkeit, beliebig viele Teilnehmer zu verwenden, allerdings kostet jeder weitere Teilnehmer Bandbreite. Schwerwiegender sind aber die Einschränkungen durch Kollisionsvermeidungen und Synchronisierungsalgorithmen. Hinzu kommt, dass es bei vielen Teilnehmern zu Beeinträchtigungen der Reaktionszeit der einzelnen Stationen kommen kann [7].

Frequency Division Multiplex (FDM)

FDM unterteilt das Frequenzband in verschiedene Kanäle. Dies bietet unter der Voraussetzung, dass die Teilnehmerzahl x kleiner gleich der Anzahl der Kanäle k liegt, und die Bandbreite der Kanäle ausreichend groß ist, gute Bedingungen für Realzeitsysteme. Sind diese Voraussetzungen nicht gegeben, muss FDM mit einem der anderen Multiplexverfahren kombiniert werden, was zur Folge hat, dass deren schlechte Eigenschaften mit übernommen werden [7].

Code Division Multiplex (CDM)

Bei CDM werden die Signale mittels eines Codierungsverfahrens verändert, so dass selbst nach Überlagerungen und kleinen Störungen die Daten vom Empfänger gelesen werden können. Dies geht aber zu Lasten der Reaktionszeit und erzeugt eine höhere Komplexität: Der En- und Decodiervorgang benötigt zusätzlich Zeit und die Stationen müssen ihre Sendestärke kontrollieren, um kein anderes Signal zu übertönen. Bei einer sehr großen Teilnehmerzahl bietet CDM die einzige effektive Lösung, deswegen wird es dort auch bei Realtime Anwendungen eingesetzt (Bsp. UMTS) [7].

3.2.2 Vergleich der MAC Frames

Bevor die Zugriffsverfahren untersucht werden, werden die MAC Frames miteinander verglichen, um festzustellen, ob es auch dort Unterschiede gibt. Auch das Verhältnis der Zusatzinformationen zur Nutzlast ist für Echtzeitapplikationen interessant. Als Vertreter für die Kabelnetzwerke wird Ethernet 802.2 verwendet. WLAN nach 802.11 steht für die kabellosen Varianten.

IEEE 802.3 MAC Frame

Der Frame (siehe Abb. 3.3) besteht aus sechs Feldern. Dabei nimmt der Header 19 Byte in Anspruch und gliedert sich in die Felder Preamble, Start Frame Delimiter (SFD), Destination und Source Adress. Jeder 802.3 Frame beginnt mit einer Folge von Einsen und

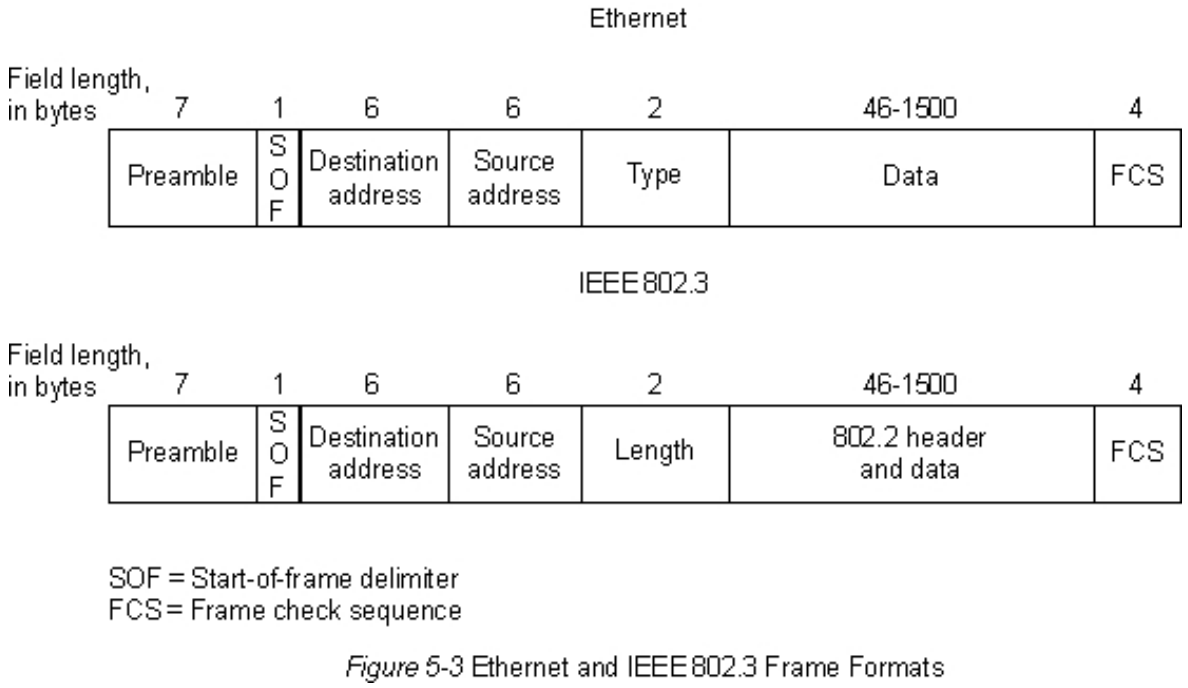

Figure 5-3 Ethernet and IEEE 802.3 Frame Formats

Abbildung 3.3: Der 802.3 MAC Frame

Nullen im Preamble Feld. Dies dient dazu, anderen Stationen mitzuteilen, dass ein Frame unterwegs ist. Außerdem gibt es den anderen Stationen Zeit, sich auf die Bitfolge zu synchronisieren. Das Feld SFD signalisiert den Beginn der Ziel Adresse. In den beiden Adressfeldern steht jeweils eine 6 Byte lange MAC Adresse. Danach kommt der variable Payload von 46 bis 1500 Byte. Das Feld FCS enthält zum Schluss noch eine CRC Prüfsumme, mit deren Hilfe beschädugte Frames erkannt werden können. Insgesamt besteht der Frame aus 24 Byte ohne das Payloadfeld.

IEEE 802.11 MAC Frame

Beim 802.11 MAC Frame 3.4 existieren insgesamt neun Felder. Der Header allein enthält bereits sieben Felder. Das Feld Frame Control FC enthält WLAN spezifische Informationen, wie zum Beispiel: Versionsnummer, WEP, Power Management. Im Feld Duration ID wird die Dauer der Übertragung angeben. Die ersten drei Adressen enthalten neben der Quell- und Zieladresse auch noch den Basis Service Set Identifyer. Das Feld Sequence Control dient zur Einordnung des Frames. Nur bei Kommunikation zwischen zwei verteilten Systemen wird auch das vierte Adressfeld genutzt. Wie bei Ethernet enthält das Feld FCS eine Prüfsumme zur Fehlererkennung.

3.2.3 MEDIA ACCESS CONTROL Protokolle in Ethernet Kabelnetzwerken

Das IEEE hat für Ethernet Netzwerke drei MAC Verfahren aufgestellt: TokenRing (802.5), TokenBus (802.4) und CSMA/CD (802.3). Alle drei Verfahren werden im folgenden Ab-

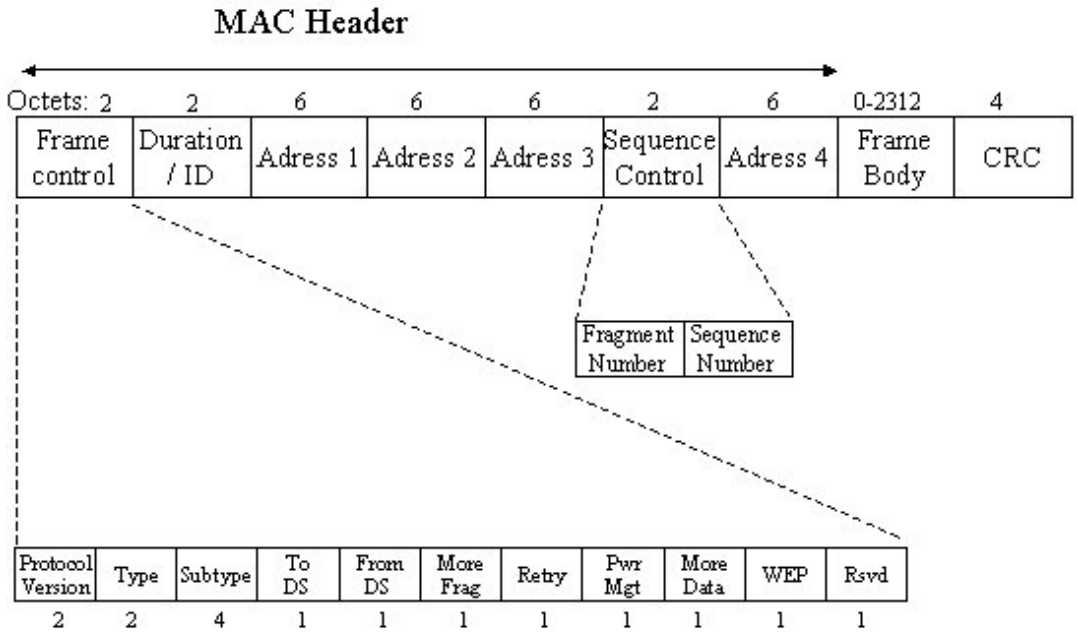

Abbildung 3.4: Der 802.11 MAC Frame

schnitt untersucht, wobei 802.3 heutzutage den mit Abstand größten Marktanteil stellt.

IEEE 802.5 Token Ring

Token Ring wurde ursprünglich von IBM entwickelt und erst später vom IEEE als Standard verabschiedet. Dabei bilden alle Teilnehmer einen geschlossen Ring. Die Übertragungsgeschwindigkeit beträgt 4 oder 16 MBit/s. Dabei lassen sich bis zu 260 bzw. 72 (bei 16 MBit/s) Teilnehmer anschließen.

Möchte ein Teilnehmer Daten übertragen, muss er auf die Zuteilung des Tokens (spez. Bitmuster) warten. Erreicht ihn das Token, so kann er mit der Datenübertragung beginnen und ein Frame übertragen. Dabei werden die Daten von einem Teilnehmer zum nächsten immer in der gleichen Richtung weitergeleitet. Erreicht den Sender der eigene Frame wieder, so erzeugt er das Token und sendet es weiter.

Das Verfahren (siehe Abb. 3.5) lässt sich mit Polling Mechanismen vergleichen, nur das bei Token Ring die Anordnung der Teilnehmer den Master ersetzt. Allerdings hängt der Zeitpunkt der erneuten Zuteilung von der Anzahl der sendewilligen Stationen ab [4].

IEEE 802.4 Token Bus

Zu Beginn muss man sagen, dass die IEEE den Standard 802.4 zurückgenommen hat. Beim Token Bus hängen die Teilnehmer an einer Leitung, die im Gegensatz zum Token Ring nicht geschlossen ist. Allerdings wird über diese Topologie ein logischer Ring gelegt.

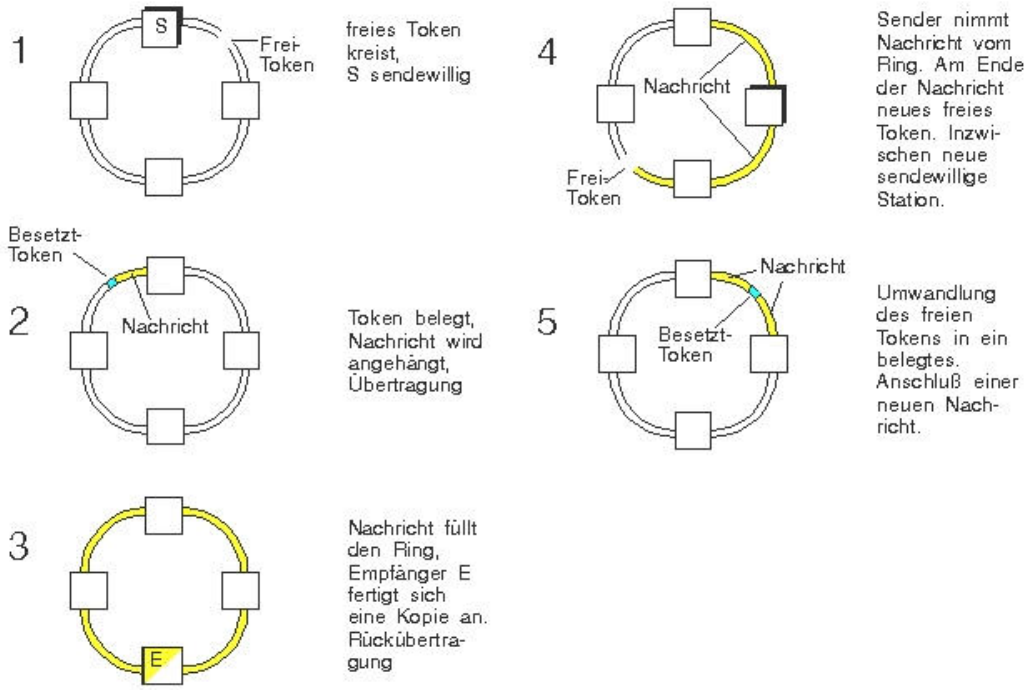

Abbildung 3.5: Das Verfahren im Überblick [4]

Jede Station kennt die Adresse der nachfolgenden Station: Es entsteht ein geschlossener Ring, da die letzte Station als Nachfolger die erste Station hat.

Im Gegensatz zum Token Ring bekommt jede Station einen festen Zeitintervall zur Übertragung. Somit kann ein Zeitpunkt t_{max} garantiert werden, zu dem eine Station spätestens wieder senden kann. Dies ist im Hinblick auf harte Realzeitsysteme interessant: Liegt t_{max} der erneuten Token Zuteilung soweit unterhalb von t_{max} der Echtzeitapplikation, dass die Datenmenge rechtzeitig übertragen werden kann, bietet Token Bus Einsatzmöglichkeiten für Echtzeitapplikationen.

Dies zeigt sich auch daran, dass Token Bus von General Motors in der Steuerung von Produktionsanlagen eingesetzt wurde. Außerdem kennt Token Bus eine Prioritätseinteilung: Wichtige Stationen können im logischen Ring bei einer Runde mehrmals das Token erhalten.

Die Übertragungsgeschwindigkeit liegt bei 1, 5 oder 10 MBit/s.

Diese Methode eignet sich auch für andere, nicht geschlossene Topologien: Es kann jederzeit ein logischer Ring aller Teilnehmer gebildet werden [5].

IEEE 802.3 Carrier Sense Multiple Access with Collision Detection (CSMA/CD)

CSMA/CD ist heute der weit verbreiteste Standard in Ethernet Netzwerken. Die Teilnehmer werden in einer Baumstruktur miteinander verbunden. Möchte eine Station Daten

senden, überprüft sie das Medium: Ist die Leitung frei, wartet sie noch mal den Interframe Gap ab, bevor sie erneut überprüft, ob noch immer eine Übertragung möglich ist. Falls ja, beginnt sie mit der Übertragung. Gleichzeitig hört sie aber die Leitung ab, um Kollisionen, die durch gleichzeitiges Beginnen einer Übertragung oder durch lange Kabelverbindungen entstehen, zu erkennen. Erkennt eine Station eine Kollision, bricht sie die Übertragung ab, und sendet stattdessen ein Jaming Signal. Danach greift ein Backoff Algorithmus, der verhindern soll, dass beide Stationen wieder zur gleichen Zeit mit der Übertragung beginnen.

Bei zwei Rechnern und voll Duplexfähigen Netzwerkkarten ist CSMA/CD ohne Einschränkungen hart echtzeitfähig. Erst bei mehreren Rechnern können Probleme durch Kollisionen auftreten. Dies kann allerdings umgangen werden, durch den Einsatz intelligenter Switches, statt der herkömmlichen Hubs: Die Switches erkennen die Ziel MAC-Adresse eines Pakets, und leiten es nur noch an den betroffenen Rechner weiter. Dies wird durch kleine Puffer erreicht, in denen die Pakete zwischengespeichert und ausgewertet werden. Dieser Vorgang kostet allerdings Zeit und führt zu einem neuen Problem: Ist der Speicher voll, werden keine neuen Pakete angenommen. Es kommt zu Paketverlusten.

IEEE 802.3 kann mittlerweile mit bis zu 1000 MBit/s als Bandbreite betrieben werden [6].

3.2.4 MEDIA ACCESS CONTROL Protokolle in Funknetzwerken

Zusätzlich zum Problem, den Zugriff auf das gemeinsame Medium zu regeln, kommen spezifische Probleme der Funknetzwerke hinzu. Während bei Kabelnetzwerken alle Teilnehmer bekannt sind, ist das bei Funknetzwerken nicht gewährleistet. Dies führt zu folgenden Schwierigkeiten:

Ein Problem, dass daraus resultiert, ist das sogenannte **Hidden Station** Problem: Möchte eine Station A der Station B Daten senden, hört es das Medium ab. Dabei kann es aber auf Grund der räumlichen Ausdehnung nicht hören, dass eine Station C (die Hidden Station) bereits begonnen hat, Daten zur Station B zu senden. Also fängt auch A an, Informationen zu übertragen: Bei der Station B werden die Informationen durch die Überlagerung der Signale zerstört (siehe Abb. 3.6).

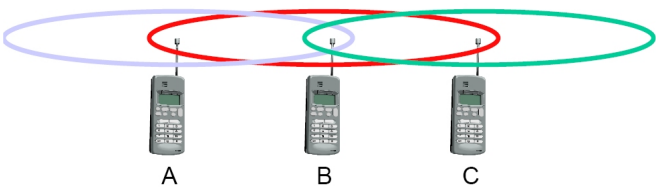

Abbildung 3.6: Das Hidden Station Problem [7]

Ein weiteres Problem schränkt die Effektivität von Funknetzwerken ein. Eine Station A möchte Informationen zur Station D senden. Bei der Überprüfung des Mediums stellt sie

fest, dass bereits die Nachbarstation B Daten zur Station C überträgt. Dabei liegen die Sationen soweit auseinander, dass eine Übertragung von A nach D problemlos möglich wäre: Die Station A ist der Station B „ausgeliefert“. Das Problem heißt **Exposed Terminal**.

(Slotted) Aloha

Aloha ist ein zufälliges TDM-Verfahren. Es wurde erstmals in einem Forschungsnetz der Universität von Hawaii verwendet. Eine Station, die Daten senden möchte, versucht dies einfach, in der Hoffnung, das Medium ist frei, und der Empfänger erhält die Daten. Glückt dieser Versuch, erhält der Sender eine Quittung des Empfängers.

Da dieser Vorgang aber nicht zentral gesteuert wird, ist die Wahrscheinlichkeit sehr hoch, dass auf Grund anderer sendender Stationen es häufig zu Kollisionen kommt. Der Empfänger kann keine Daten mehr aus dem empfangen Signal entnehmen. Des weiteren kann es zum Einen passieren, dass eine Station so ungünstig den Übertragungszeitraum wählt, dass sie sowohl das Ende einer fast fertigen Übertragung, als auch den Anfang einer neuen Übertragung überlagert. Dadurch sind alle drei Übertragungen nicht erfolgreich.

Um das zu verhindern, wurden später feste Zeitschlitze(sog. Slots) eingeführt, in denen eine Station senden darf. Daher auch der Name der Erweiterung: Slotted Aloha.

Da Aloha ursprünglich mit nur drei Station betrieben wurde, bestand keine Notwendigkeit, die Probleme von Hidden Station und Exposed Terminal zu behandeln.

Insgesamt haben Aloha und Slotted Aloha bei einem Poissonverteilten Datenaufkommen jedoch nur eine Wirksamkeit von 16 Prozent (bzw. 32 Prozent). Da Kollisionen sowohl beim Senden der Daten als auch beim Verschicken der Empfangsquittung entstehen können, ist (Slotted) Aloha denkbar ungeeignet für Realtime Applikationen (Abb. 3.7) [7].

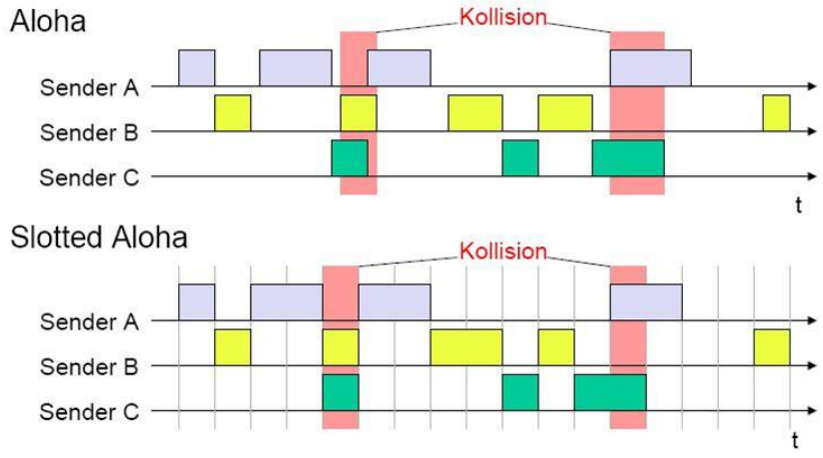

Abbildung 3.7: Schema von Aloha, bzw. Slotted Aloha [7]

Demand Assigned Multiple Access (DAMA)

Um der schlechten Ausnutzung von (Slotted) Aloha zu begegnen, wurde DAMA entwickelt. Dabei werden vor Beginn des Übertragungszeitraums spätere Timeslots reserviert. Die Reservierung erfolgt entweder mittels Slotted Aloha oder eines festen Zeitmultiplex.

Bei der Slotted Alhoa Version nimmt man in Kauf, dass es zu Kollisionen kommt. Dafür ist dort eine dynamische Teilnehmerzahl möglich. Bei einem festen Zeitmultiplex hingegen, muss entweder die max. Teilnehmerzahl bekannt sein, oder es gibt eine zentrale Station, die den Zeitmultiplex entsprechend der aktuellen Teilnehmerzahl anpasst.

Das Problem ist, dass dafür zum einen ein Synchronisierungsmechanismus nötig ist, zum anderen muss jeder Teilnehmer eine Reservierungsliste führen. Für Realtimeanwendungen mit sehr kleinem t_{max} kommt hinzu, dass der Reservierungsvorgang unnötig Zeit kostet [7].

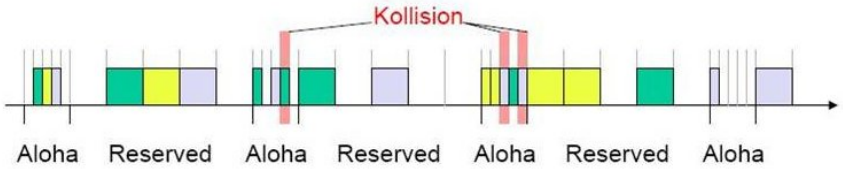

Abbildung 3.8: DAMA mit Slotted Aloha zur Reservierung [7]

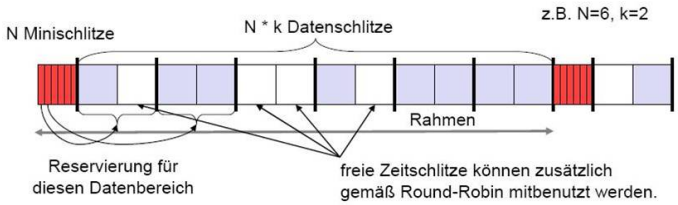

Abbildung 3.9: DAMA mit implizierter Reservierung [7]

Polling

Beim Polling kann eine Station, der Master, alle anderen Stationen erreichen. Dies bedeutet insbesondere, es existiert kein Hidden Terminal. Dann gibt es verschiedene Möglichkeiten, dass Medium aufzuteilen: Es gibt einen festen Zeitmultiplex, bei dem der Master den einzelnen Slaves nach und nach Zugriff aufs Medium gewährt, indem er sie zum Senden auffordert. Der große Vorteil besteht darin, dass es ‚solange es zu keinen Störungen von Außen kommt, Kollisionen vermieden werden können. Des weiteren kann man sogar Garantien abgeben, nach welcher Zeit eine Station wieder senden kann.

Eine weitere Möglichkeit ist, dass nach einem Startping alle sendewilligen Stationen, mittels FDM oder CDM gleichzeitig eine Zufallszahl übertragen. Die Slaves werden dann chronologisch ihrer Zufallszahl nacheinander abgearbeitet. Allerdings kann es dabei zu

Kollisionen kommen. Dadurch verliert aber das Polling seinen Vorteil gegenüber anderen Verfahren.

Durch die Vorraussetzung, dass eine Station bereits alle Stationen erreichen kann und sie selber an die anderen Stationen die Sendeberechtigung verteilt, kann weder das Hidden Station Problem noch ein Exposed Terminal auftreten. Damit bietet Polling eine Möglichkeit für eine begrenzte, und bekannte Teilnehmerzahl echtzeitigfähig zu sein [7].

Multiple Access with Collision Avoidance (MACA)

Eine Variante (Carrier Sense Multiple Access with Collision Avoidance, kurz CSMA/CA) des MACA ist zur Zeit das gängigste Verfahren für WLANs, denn sie wird vom IEEE Standard 802.11 verwendet. Dabei verwenden alle Stationen kurze Signalisierungsmitteilungen (RTS) um den anderen Stationen den Sendewunsch mitzuteilen, währenddessen warten sie auf Rückmeldungen (CTS) des Empfängers, ob er überhaupt empfangsbereit ist.

Die Mitteilungen enthalten neben der Ziel- und der Quelladresse noch die Paketlänge, damit die Übertragungsdauer abgeschätzt werden kann. Dadurch werden Probleme, wie das Hidden und Exposed Terminal, umgangen. Um Kollisionen bei den Mitteilungen zu vermeiden, werden Interframe Spaces (IFS) in Verbindung mit einem zufälligen Backoff Mechanismus in einem Wettbewerbsfenster zwischen den einzelnen Übertragungen verwendet. Dabei werden drei Arten von Abständen unterschieden:

Short IFS Muss eine Station auf eine Signalisierungsmitteilungen antworten, wartet sie nur einen kurzen Zeitpunkt ab, damit keine andere Übertragung vorher beginnen kann.

PIFS Möchte eine Station Übertragungen mit höherer Priorität versenden, wartet sie einen Zeitraum ab, der länger als der Short IFS ist, aber noch vor den normalen Übertragungen liegt. Beispiele hierfür sind Realtime Daten.

DIFS Für die normale Datenübertragung wird am längsten gewartet, damit unwichtige Übertragungen nicht unnötig das Netzwerk blockieren.

Der Backoff Mechanismus dient dazu, dass nicht alle sendewilligen Stationen gleichzeitig nach Abwarten des DIFS zu Senden beginnen. Jede Station ermittelt eine Zufallszahl, die sie zusätzlich nach dem DIFS wartet. Dabei hört sie nebenbei das Medium ab, um festzustellen, ob eine andere Station mit einer niedrigeren Zufallszahl schon zu senden begonnen hat. Bei gleicher Zufallszahl kann es natürlich zu Kollisionen kommen.

Als Realtimeumgebung bietet es statistisch gesehen eine relativ gute Vorraussetzung für weiche Echtzeit Applikationen, obwohl natürlich die Signalisierungsmitteilungen und das Wettbewerbsfenster für Verzögerungen und Overhead sorgen. Mit steigender Teilnehmerzahl sinkt allerdings die Qualität erheblich. Für harte Echtzeitbedingungen bietet MACA keine Möglichkeit, da ein erfolgreicher Mediumzugriff in einer bestimmten Zeit nicht festgelegt werden kann [7].

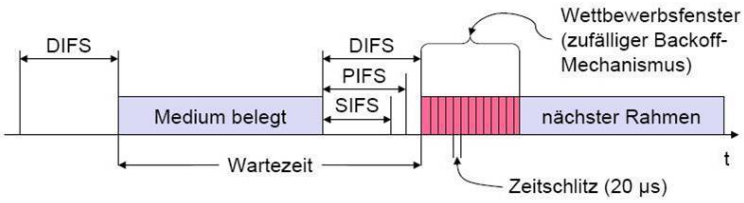

Abbildung 3.10: Die unterschiedlichen IFS bei MACA [7]

3.2.5 Zusammenfassung der Mediumzugriffsverfahren

Bei Kabelnetzwerken und Hardrealtimeanwendungen bieten die Token Systeme, insbesondere Token Bus gute Voraussetzungen. Durch den logischen Ring können Garantien für den spätesten Zeitpunkt des erneuten Mediumzugriffs abgegeben werden. Allerdings nimmt man dafür eine aufwendige Verwaltung bei Neuaufnahmen von Ringmitgliedern in Kauf.

Möchte man bei harten Echtzeitapplikationen CSMA/CD verwenden, sollte man auf die Umgebung Wert legen: Unnötige Rechner aus dem Netzwerk entfernen. Router und Switches mit ausreichend großen Puffern ausstatten, um einen Paketverlust vorzubeugen. Unter diesen Umständen kann man auch unter CSMA/CD zumindest statistische Garantien abgeben.

Auf der Wireless Seite stellt die Störanfälligkeit von Außen ein großes Problem dar. Nichts desto trotz bietet das Pollingverfahren akzeptable Voraussetzung für nahezu Hardrealtime Applikationen: Die Reaktionszeit ist zwar hoch und steigt mit der Anzahl der Teilnehmer, dafür existiert aber eine Schranke für die maximale Dauer bis zum nächsten Mediumzugriff einer Station. Allerdings geht das nur unter der Nebenbedingung, dass der Master alle Slaves in einer störungsfreien Umgebung erreichen kann.

(Slotted) Aloha ist auf Grund der hohen Ineffektivität weder für Soft- noch für Hardrealtimesysteme geeignet. Dies wird auch nicht durch den Vorteil der nicht vorhandenen Verwaltung von Aloha wieder kompensiert.

Wenn bei DAMA Slotted Aloha zur Reservierung verwendet wird, kann es passieren, dass zwei oder mehrere Stationen keinen Timeslot reservieren können: Es kann keine obere Schranke t_{max} definiert werden. Für Softrealtime Applikationen kommt hinzu, dass das Reservierungsverfahren unnötig Zeit kostet.

MACA bietet für viele Teilnehmer und weiche Echtzeitbedingungen eine Alternative zum Polling. Allerdings wird dadurch Overhead in Kauf genommen. Für Hardrealtime ist es ungeeignet, da durch den zufälligen Backoff Mechanismus keine obere Schranke definiert werden kann.

3.3 Realtimeprotokolle

Das nächste Kapitel beschäftigt sich mit den Realtimeprotokollen. Dabei werden die Protokolle im OSI/ISO Referenzstack eingeordnet und später insbesondere auf ihre Eignung in WLANs geprüft.

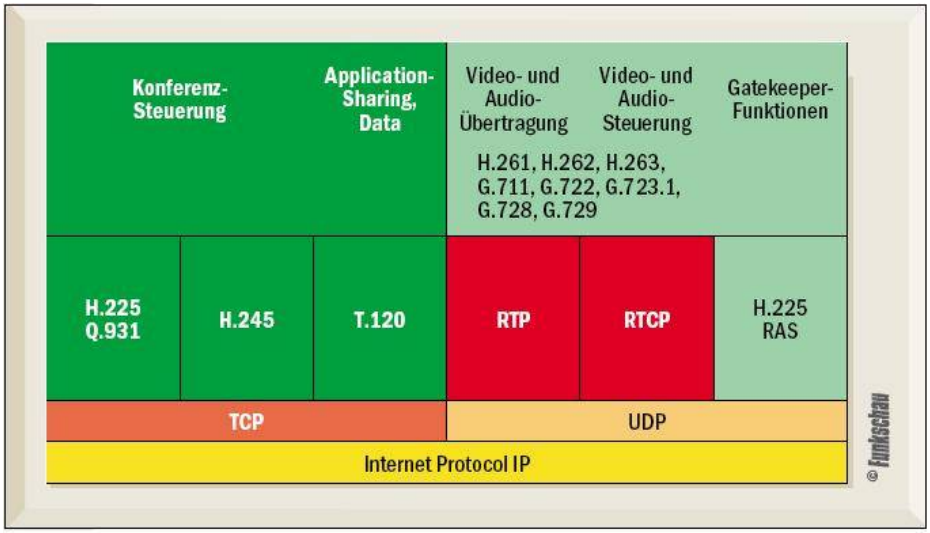

Abbildung 3.11: RTP und RTCP eingeordnet im TCP/IP Stack [11]

3.3.1 Realtime Protocol (RTP nach IETF RFC3550)

Das Realtime Protocol setzt direkt auf das User Datagram Protocol (UDP) auf. Nach dem ISO/OSI Referenzmodell ist es in der Ebene 7 / Applicationlayer einzuordnen, obwohl es auch Meinungen gibt, die RTP noch in die Ebene 4 / Transportlayer einordnen.

Es ist kein vollständig spezifiziertes Protokoll, sondern stellt wichtige Funktionen und einen Rahmen für Softrealtime Umgebungen zur Verfügung. Da es auf UDP aufsetzt, gibt es keine direkte Rückmeldung über Paketverluste. Es existiert auch kein Mechanismus zur erneuten Anforderung von IP-Paketen. Da RTP in erster Linie für Audio- und Videoübertragungen konzipiert wurde, werden die Verluste in Kauf genommen, da sie lediglich Qualitätseinbußen produzieren.

Der RTP-Header enthält ein Feld für die Versionsnummer, da mittlerweile die Version 2.0 (RFC 3550) die erste Version (RFC 1889) abgelöst hat. Da es für manche Anwendungen ein Vorteil ist, Pakete fester Größe zu verarbeiten, gibt es ein Paddingfeld, indem die Anzahl der aufgefüllten Paddingbits steht. Des weiteren enthält jeder Header eine Sequenznummer, mit deren Hilfe die Reihenfolge und die Eindeutigkeit von Paketen festgestellt werden kann. Darüber hinaus werden die Pakete mit einem Zeitstempel versehen, damit Applikation in der Lage sind, die Inhalte zu synchronisieren. Mit dem Synchronisation Source Identifier verfügt der RTP Header über ein Feld, in dem die Datenquelle identifiziert werden kann. Da RTP möglichst flexibel gehalten wurde, existiert keine eindeutige

Beschreibung für den Payload, sondern es existieren verschiedene Profile für einzelne Anwendungsgebiete. Als Beispiel sei hier die RFC 2038 für MPEG1 / 2 Video genannt. Damit eignet sich RTP ausschließlich, aber dafür sehr gut für Softrealtime Anwendungen [8].

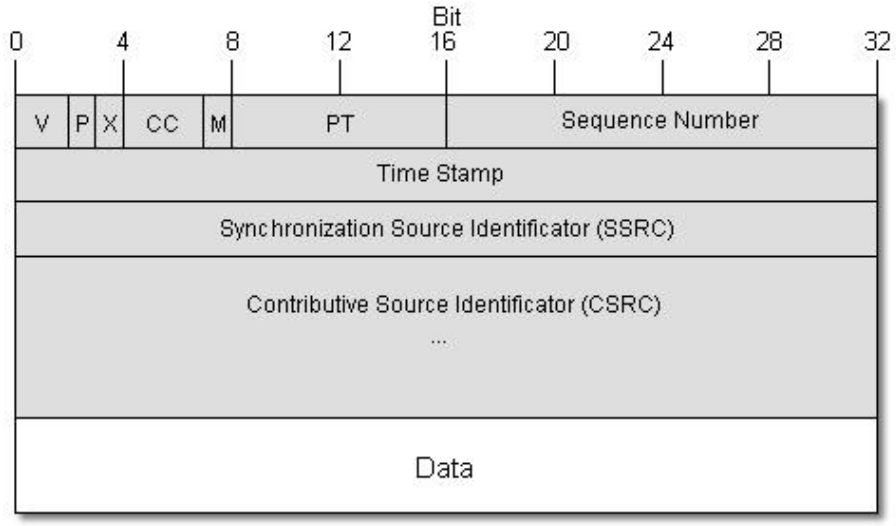

Abbildung 3.12: Der RTP Frame [12]

3.3.2 Realtime Transport Control Protocol (RTCP nach IETF RFC3550)

Das RTCP ist kein eigenständiges Protokoll, sondern eine Erweiterung zu RTP. Es setzt ebenfalls auf dem UDP Protokoll auf. Für die Datenübertragung wird weiterhin RTP verwendet. Zusätzlich aber werden kurze Mitteilungsnachrichten verschickt, um Qualitätsaussagen über die einzelnen Verbindungen zu treffen.

Dabei senden alle Sitzungsteilnehmer regelmäßig Nachrichten, so dass die Sendestation eine Übersicht hat, welche Verbindungsqualität jeder einzelne Teilnehmer hat. Daraufhin kann die Sendestation die Kompressionsraten der Daten anpassen. Dadurch werden die einzelnen Teilnehmer mit der optimalen Datenmenge versorgt. Darüber hinaus können noch zusätzliche Informationen über die Teilnehmer verschickt werden, wie zum Beispiel Name, eMail Adresse usw. Auch mit dem RTCP bleibt RTP nur für Softrealtime Umgebungen attraktiv [8].

3.3.3 RTnet der Universität Hannover

RTnet wurde als Feldbus Ersatz mit Ethernet Komponenten konzipiert. Dabei wird der herkömmliche ISO/OSI Referenzstack schon ab der MAC Schicht verändert. Um die Probleme von CSMA/CD zu umgehen, stellt die neue RTMAC Schicht zwei Verfahren zur Verfügung: Ein Tokenpassing-Verfahren, bei dem mittels des Token die sendeberechtigte Station ermittelt wird. Das zweite Verfahren ist ein fester Zeitmultiplex. Beide Verfahren erfordern einen Master, der entweder das Token an die richtige Station weitergibt, oder

das Auf- und Verteilen der Timeslots bei TDMA übernimmt. Als zusätzliche Erweiterung wird das dynamische ARP Protokoll durch eine statische Variante ersetzt.

Über der RTMAC Schicht setzt RTnet auf UDP als Protokoll. Darüber liegen dann direkt die Echtzeitapplikationen, die direkt auf RTnet zugreifen können. Da RTnet davon ausgeht, dass im Netzwerk lediglich RTnet Rechner sind, von denen ja immer nur eine Station sendet, wird der Einsatz von Hubs und nicht von Switches empfohlen.

Um in diesem Netzwerk aber auch gewöhnlichen TCP/IP Verkehr zu ermöglichen, bietet RTnet eine zusätzliche virtuelle Interface Schnittstelle an. Der dort aufkommende Traffic wird dann von RTnet getunnelt.

Auf Grund dieser gravierenden Eingriffe in den TCP/IP Stack wird für jeden Netzwerkkartenchip eine eigene Implementierung des Netzwerktreibers nötig. Obwohl für viele Kabelnetzwerkkarten Treiber erstellt wurden, existieren noch keine Implementierungen für Wireless-Adapter. Bisher ist RTnet eine Linuxlösung, die sich aber auf andere Echtzeitbetriebssysteme portieren lassen soll. Die Realtime Eigenschaften für Hardrealtime hat RTnet bereits in Versuchen unter Beweis gestellt. Dabei ließe sich RTnet sicher auch für Softrealtime Umgebungen einsetzen [9].

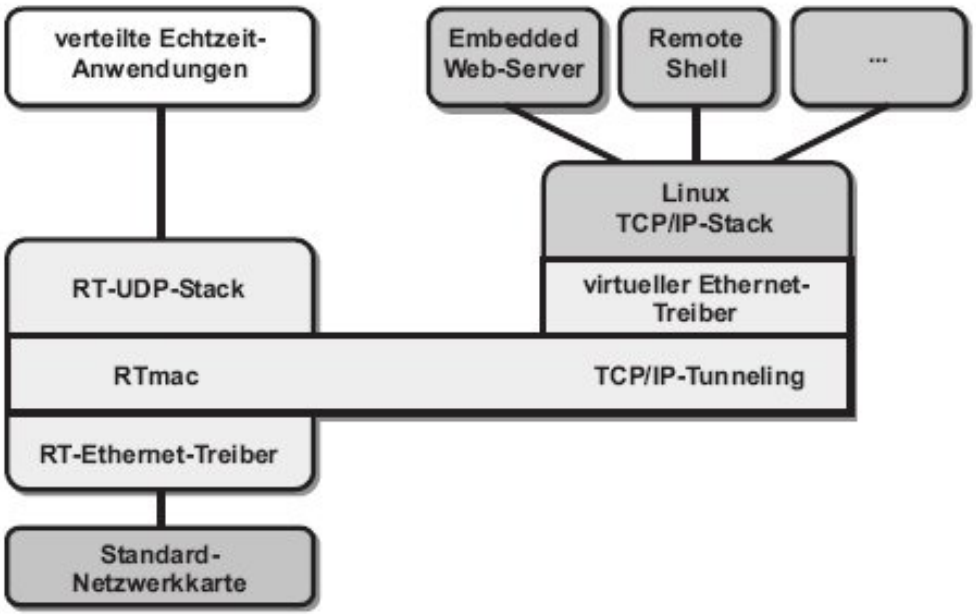

Abbildung 3.13: Der Aufbau von RTnet [9]

3.3.4 Protokolle Wireless tauglich?

RTP bzw. die Kombination RTCP setzt auf dem User Datagramm Protokoll auf. Da UDP auf Schicht 4 aufsetzt, spielt das eigentliche Medium keine Rolle. Bei der Kombination von RTP/RTCP werden durch die Berichte der einzelnen Teilnehmer optimale Datenraten

erzielt. Dadurch können sogar kurzfristige Beeinträchtigungen der Übertragungsqualität von mobilen Systemen ausgeglichen werden.

Des weiteren besitzt UDP keine kontraproduktiven Mechanismen, wie TCP. Im Wireless Fall werden Paketverluste durch Fehler bei der Übertragung von normalen TCP als Stau interpretiert. Daraufhin setzt TCP die Datenrate drastisch herunter, obwohl es unnötig ist. Es gibt zwar Konzepte für mobile TCP, aber dort hat sich kein Standard durchgesetzt. RTnet hingegen ist auf Grund seiner Eingriffe bis in die Schicht 2 auf eigene Treiber angewiesen. Bisher wurde RTnet nur für herkömmliche Kabelnetzwerke konzipiert, weshalb es in der aktuellen Version nicht auf WLANs übertragbar ist.

3.3.5 Kommerzielle Realtime Protokolle

Die Firma Hilf! hat ein Protokoll namens Realwire entwickelt. Die verfügbaren Informationen besagen, dass es auf UDP aufsetzt und über einen festgelegten Zeitmultiplex verfügt. Genauere Informationen liegen leider nicht vor.

Von FSMLabs gibt es eine Variante von RTnet mit dem Namen LNet, dass ebenfalls für RTLinux konzipiert wurde. LNet soll den Applikationen direkt zur Verfügung stehen ohne Umwege über das Betriebssystem gehen zu müssen. Dabei fängt es wie RTnet die Pakete direkt nach nach der Schicht 2 ab. Wie dies realisiert wird, ist leider nicht beschrieben.

Die kommerziellen Lösung bieten zwar den Vorteil der Produktbetreuung, und evtl. der Anpassung an persönliche Bedürfnisse. Dafür legen sie ihre tatsächliche Funktionsweise nicht offen, sondern beschreiben lediglich die Eigenschaften.

3.4 Existierende Systeme

3.4.1 Roboter im Kaufhaus (RoBoKa)

RoBoKa ist ein Projekt des Lehrstuhls für Rechnerarchitektur der Universität Rostock. Dabei wurde ein Roboter gebaut, der mittels WLAN nach IEEE 802.11 gesteuert werden sollte. Der Roboter konnte nachts im Kaufhof von Internetusern über ein Java Applet gesteuert werden. Dabei sendete der Roboter bis zu 30 Bilder pro Sekunde. Gleichzeitig erhielt er Befehle, wie zum Beispiel vorwärts, links drehen, rechts drehen.

Was auf den ersten Blick wie ein hartes Echtzeitsystem aussieht (Kollisionsgefahr des Roboters mit anderen Objekten), wurde von den Verantwortlichen entschärft, indem der Roboter rundum mit Ultraschallsensoren ausgestattet wurde. Dadurch hielt der Roboter selbstständig vor Hindernissen, anstatt auf den Stoppbefehl zu warten [10].

Abbildung 3.14: Foto vom Roboter im Kaufhaus [10]

3.5 Zusammenfassung

3.5.1 Fazit

Im Gegensatz zu Token Ring, das Hardware und Topologie abhängig ist, lässt sich der logische Ring von Token Bus auch auf andere Systeme übertragen. Darüber hinaus bietet Token Ring keine Möglichkeit, den maximalen Zeitpunkt der erneuten Mediumzuteilung zu bestimmen. Damit bietet Token Ring für harte Realtimenetzwerke keine Lösung. Da TokenBus einen maximalen Zeitpunkt bis zur erneuten Zuteilung der Sendeberechtigung besitzt, lässt es sich problemlos auch für Hardrealtime Anwendungen einsetzten. Dafür wurde es ursprünglich auch von GM konzipiert.

Netzwerke mit CSMA/CD lassen sich selbst mit Aufwand in der Konzeption und im Aufbau nur bedingt für Realtime Umgebungen einsetzen. Während man bei kleinen Datenmengen und einer geringen Teilnehmerzahl noch sehr gute Ergebnisse erzielt, steigt mit größeren Datenaufkommen und höherer Teilnehmerzahl deutlich das Risko, dass Zeitpunkte nicht mehr eingehalten werden können. Allerdings bieten sie für heutige Softrealtime Applikationen mehr als genügend Bandbreite und schnelle Reaktionszeiten für die Bedüfnisse einer begrenzten Benutzerzahl.

Die Zugriffsverfahren bei WLAN besitzen alle Einschränkungen: Entweder sie versuchen die wireless typischen Probleme zu behandeln, oder sie machen Vorgaben, unter welchen Bedingungen das Verfahren funktioniert. Zu alledem kommt noch die Störanfälligkeit von WLANs, so dass sie unter keinen Umständen für kritische Echtzeitsysteme geeignet

sind. Mit Polling gibt es zwar ein Verfahren, dass Zeitgarantien abgeben kann, allerdings stets unter der Voraussetzung, dass es keine äusseren Störquellen gibt. Für Softrealtime Umgebungen sind bis auf (Slotted) Aloha alle Verfahren geeignet, da mittlerweile für WLANs ausreichend Bandbreite zur Verfügung stehen.

Bei den Protokollen erkennt man in den zwei vorgestellten Verfahren, je einen Vertreter für Soft- bzw. Hardrealtime. Während RTP/RTCP für Audio/Videodaten optimiert wurde, haben die Entwickler von RTnet insbesondere auf deterministische Zeitaussagen Wert gelegt. Dieser Unterschied zeigt sich auch deutlich an den Modifizierungen des OSI/ISO Referenzstacks: RTP/RTCP setzt auf UDP auf, ohne an den restlichen Schichten Änderungen vorzunehmen. Dadurch spielt auch das Übertragungsmedium keine Rolle. Durch die regelmäßigen Sende- und Empfangsberichte können sogar leichte Schwankungen von Wireless Verbindungen durch Anpassung der Datenraten ausgeglichen werden. RTP/RTCP sind freie Standards für das Internet.

Bei RTnet wird der ISO/OSI Referenzstack bereits ab der Schicht 2 modifiziert, was spezielle Treiber erfordert. Dafür kann RTnet Zeitgarantien abgeben und ist darüber hinaus als kostenlose Implementierung erhältlich. Zudem wird RTnet ständig weiterentwickelt, was die bereits jetzt erzielten Eigenschaften verbessert, bzw. RTnet um weitere Eigenschaften erweitert.

3.5.2 Ausblick

In Zukunft werden vor allem Softrealtime Applikationen immer mehr Einzug in Computersysteme halten. Insbesondere Audio- und Videoübertragungen, aber auch Computerspiele werden die Netzwerke der Zukunft belasten. Dabei hat sich aber schon gezeigt, dass die aktuellen Systeme, egal ob mit Kabel oder ohne, ausreichend für die Beürfnisse der meisten heutigen Nutzer sind. Daher werden zukünftig die Datenraten in Verbindung mit den steigenden Anforderungen der User weiter steigen. Getreu dem Motto: „Warum soll ich mich mit 500KBits zufrieden geben, wenn ich 2MBits haben kann?“

Im Bereich der Mediumzugriffsverfahren werden keine Revolutionen mehr erwartet, sondern nur noch Optimierungen der vorhandenen Methoden. Durch bessere Hardware Komponenten können Sicherheitsabstände zwischen einzelnen Übertragungen weiter reduziert werden. Zusätzlich ermöglichen bessere Komponenten kürzere Synchronisierungsabschnitte bei WLANs. Dabei werden die Reaktionszeiten von Funknetzwerken hinter den Werten der Kabellösungen liegen bleiben. Dies liegt an den charakteristischen Merkmalen des kabellosen Mediums. Die Auswirkungen lassen sich zwar minimieren, aber auch in Zukunft nicht beseitigen.

Auf Grund der großen Verbreitung und den kostengünstigen Hardwarekomponenten gibt es immer mehr Projekte, die versuchen, Protokolle wie RTnet zu entwickeln, damit Ethernet als Ersatz zu teuren Bussystemen verwendet werden kann. Sie basieren aber in der Regel darauf, dass man einen stark verkürzten OSI/ISO Referenzstack verwendet, und auf verbindungslose Protokolle, wie zum Beispiel UDP setzt.

Die Technik der Zukunft wird die Echtzeiteigenschaften der Netzwerktechniken immer weiter verbessern: Kürzere Reaktionszeiten, mehr Bandbreite, Garantien für Mediumzugriffe, usw. Aber das ideale Netzwerk wird man nicht erreichen können.

Literaturverzeichnis

[1] MONITOR, Voice over IP, Ausgabe 11/2001,
http://www.monitor.co.at/index.cfm?storyid=4180

[2] Prof. Thomas Huckle, Softwarefehler und ihre Folgen, Vortrag 2.12.1999,
http://www5.in.tum.de/~huckle/bugs.html

[3] Zweites Deutsches Fernsehen, Das digitale Schlachtfeld, Dokumentation 23.03.2004,
http://www.zdf.de/ZDFde/inhalt/22/0,1872,2112790,00.html

[4] SIEMENS AG, Das Siemens Online Lexikon, 03/2004,
http://www.networks.siemens.de/solutionprovider
/_online_lexikon/2/f001592.htm

[5] Prof. Dr. W. Kowalk, Vorlesung Rechnernetze, 19.03.2004,
http://einstein.informatik.uni-oldenburg.de/rechnernetze/tokenbus.htm

[6] Eckhardt Stasch, Ethernet, Sommer Semester 1997,
http://www.tu-chemnitz.de/informatik/RA/kompendium/vortraege_97
/ethernet/technologie.html

[7] Prof. Jochen Schiller, Skript Mobilkommunikation, Sommer Semester 2002,
http://www.inf.fu-berlin.de/inst/ag-tech/resources/MC_material.htm

[8] Heiko Gierer, Streaming Media, Seminar 2000,
http://www.it.fht-esslingen.de/~schmidt/vorlesungen/mm/seminar
/ss00/HTML/node103.html

[9] Kiszka, Hagge, Hohmann, Wagner, RTnet - Open Source Lösung für Echtzeitkommunikation
http://www.rts.uni-hannover.de/mitarbeiter/kiszka/Kiszka03-Telematik.pdf

[10] Heiko Kopp, Heiko Buchholz, Peter Eschholz, Projekt RoBoKa, 2002
http://www.roboka.uni-rostock.de

[11] Robert Schoblick, Funkschau Heft 26, 2000
http://www.funkschau.de/heftarchiv/pdf/2000/fs26/fs0026047.pdf

[12] Unbekannt, Internetseite zur Funktionsweise von Breitband Netzwerken
http://www.breitband-isdn.ch/technic/ip/

Kapitel 4

Micro-Mobility in IP-based Networks

Philipp Appelhoff

In der unmittelbaren Vergangenheit hat die IETF Mobile IP Working Group eine Reihe von Erweiterungen des bekannten Mobile IP Protokolls behandelt, das Mitte der 90er Jahre entwickelt wurde, um der in naher Zukunft radikal wachsenden Zahl von sogenannten Wireless Subscribern, also Nutzern mobiler Datenkommunikationsdienste, gerecht zu werden und ihnen die gleichen Möglichkeiten zu bieten, die sie aus drahtgebundenen Netzwerken kennen. Bekanntlich ergeben sich mit wachsender Mobilität und der immer schnelleren Verbreitung kabelloser Kommunikationssysteme diverse Herausforderungen, denen IP-basierte Systeme aufgrund des Konzepts der hierarchischen Adressierung nicht ohne weiteres gerecht werden können. Mit dem Mobile IP Protokoll wurde eine Lösung vorgestellt, die es dem Nutzer ermöglicht, an einem beliebigen Ort mit Zugang zum Internet auch unter seiner alten Adresse erreichbar zu sein. Steigt jedoch die Geschwindigkeit der Fortbewegung und damit die Frequenz der zu erwartenden Zellenwechsel, wird die mangelnde Effizienz des Protokolls deutlich. Micro-Mobility Protokolle erweitern daher den Ansatz von Mobile IP, um das Basisprotokoll mit Hilfe von Handoff-Optimierung und Paging, mit der Strukturierung von Netzen und effizienteren Sicherheitskonzepten zu verbessern. Diese Arbeit stellt dabei kurz die Probleme von Mobile IP heraus, gibt anschliessend einen Überblick über verschiedene Ansätze, um anschließend einige bekannte und charakteristische Protokolle und Konzepte (Fast und Proactive Handoff, Hierarchical Mobile IP, Cellular IP und Hawaii) vorzustellen. Mit Hilfe der Columbia Micro-Mobility Suite (CMS) [17] werden die drei zuletzt genannten Protokolle auf Basis bestehender Implementierungen verglichen, um die gewählten Ansätze schließlich zu bewerten.

Inhaltsverzeichnis

4.1 Einleitung 79
4.2 Micro-Mobility vs. Macro-Mobility 80
4.2.1 Mobile IP 80
4.2.2 Probleme von Mobile IP 81
4.2.3 Definition von Micro-Mobility 82
4.3 Optimierungsansätze von Micro-Mobility 83
4.3.1 Schnelle Handoff-Erkennung 83
4.3.2 Hierarchical Mobility 84
4.3.3 Paging 86
4.3.4 Fast Security 86
4.3.5 Das Triangular Routing Problem 87
4.4 Die bekanntesten Protokolle 88
4.4.1 Hierarchical Mobile IP 88
4.4.2 Fast Handoff und Proactive Handoff 89
4.4.3 Cellular IP 90
4.4.4 Hawaii 93
4.5 Vergleich der Protokolle 96
4.5.1 Konzepte 96
4.5.2 Simulation 97
4.6 Fazit 100

4.1 Einleitung

Es ist nicht mehr zu übersehen, daß drahtlose Kommunikationsgeräte mittlerweile auch den Bereich der IP-basierten Netze erreicht haben. Bereits im Jahre 2000 waren 20% aller verkauften PC´s tragbar [19]. Durch die Medien kennt bereits heute jeder Begriffe wie WLAN, Bluetooth oder auch UMTS. Mobilität ist zum Standard geworden, hinter dem jedoch viel mehr steckt als nur auf Kabel zu verzichten, oder wie es die Werbung eines großen deutschen Kommunikationsanbieter ausdrückt, kabellos online zu sein. Dennoch erwartet der Konsument von den neuen Technologien gleiche Leistungen, wie er sie mittlerweile von seinem Desktop-PC kennt. Mit dem Internet fest verkabelt lädt er Musik aus dem Netz oder hört einen New Yorker Radiosender per Internet. Zudem wird die Zukunft, vielleicht mit der Etablierung von „Voice over IP“ als konkurrenzfähiges Telekommunikationsmedium, ganz neue Erwartungen an bestehende Netzwerke stellen.

Innerhalb von IP-basierten Netzen werden IP-Adressen jedoch dazu genutzt, um sowohl den MN als auch dessen Position im Netz eindeutig zu bestimmen. Offensichtlich ergeben sich ganz neue Probleme und Herausforderungen, wenn Geräte in solchen IP-basierten Netzen mobil werden, denn dann ist es nicht mehr möglich, die Position des Empfängers ohne weiteres über dessen IP-Adresse zu bestimmen. Gleichzeitig sehen IP-basierte Netze den Wechsel der Adresse nicht vor, da Mobilität bei der Entwicklung des Standards nicht berücksichtigt wurde. Das erste Protokoll, das sich mit diesen Problemen befasste, war Mobile IP. Dieses Protokoll wird als erstes in dieser Seminararbeit vorgestellt, um eine kurze Einführung in die Funktionsweise mobilitätsorientierter Protokolle zu geben. Danach sollen die Grenzen des Protokolls aufgezeigt werden, die insbesondere in der aufwendigen Behandlung von Zellenwechseln begründet sind. Um den Problemen zu begegnen, wurde der Begriff von Mikromobilität eingeführt, der Mobilität in enger begrenzten Bereichen behandeln soll und auf quasi nahtlose Übergänge zwischen verschiedenen Netzwerkzugangspunkten ausgerichtet ist. Dazu wurden verschiedene Optimierungsansätze herausgearbeitet, um Probleme des Mobile IP Protokolls zu beseitigen oder deren Auswirkungen zu minimieren. Nachdem diese Ansätze erläutert wurden, werden im Kapitel 4.4 verschiedene charakteristische Protokolle beschrieben. Diese stellen nur einen kleinen Ausschnitt aus bestehenden Ansätzen dar. Während die Protokolle Fast und Proactive Handoff eher als Konzeptstudien zu sehen sind, werden mit Hierarchical Mobile IP, Cellular IP und Hawaii drei Protokolle vorgestellt, die basierend auf bestehenden Implementierungen im Abschnitt 4.5 verglichen werden sollen. Abschliessend werden dann die verschiedenen Ansätze bewertet und verglichen und einige Probleme angesprochen.

4.2 Micro-Mobility vs. Macro-Mobility

4.2.1 Mobile IP

Mobile IP ist das wohl bekannteste Konzept, das sich mit den Herausforderungen von Mobilität in IP-basierten Netzwerken befasst. Mobile Kommunikationsgeräte, im Folgenden als MN - Mobile Nodes bezeichnet, werfen Probleme auf, die mit bestehenden Systemen nicht zu bewältigen sind. Wenn sich ein MN in solchen IP-basierten Netzen bewegen soll, muss sich folgerichtig auch seine IP-Adresse ändern und ankommende Pakete umgeleitet werden. Bei der Entwicklung IP-basierter System wurden solche Aspekte jedoch nicht berücksichtigt, sodaß Erweiterungen diesen Rechnung tragen müssen. Dieser Aufgabe widmete sich die Mobile IP Working Group erstmalig mit der Entwicklung des Mobile IP Protokolls, das folgenden Kriterien genügen sollte:

- Erweiterung des statischen IP-Konzeptes um Mobilität
- Erreichbarkeit von mobilen Kommunikationspartnern ausserhalb ihrer Netze
- Kompatibilität mit bestehenden Systemen und Transparenz
- einfache Erweiterbarkeit
- Sicherheit und Effizienz

Die Tatsache, daß gerade die Effizienz nicht umgesetzt wurde, bildet die Grundlage dieser Arbeit. Aber zuvor soll die grundsätzliche Funktionsweise hier vorgestellt werden. Mobile IP erlaubt es einem MN, seinen WIPPOA (Wireless IP Point of Attachment, nach [4]) beliebig zu wechseln. Dieser ändert damit zum einen den Anschluss an das Netzwerk selbst, indem er sich zum Beispiel einem neuen WLAN-Accesspoint anschließt, zum anderen muss er die eindeutige Kennung innerhalb dieses Netzes, seine IP-Adresse, ändern. Für das Angebot einer solchen neuen Adresse ist ein sogenannter Foreign Agent (FA) zuständig, der im jeweiligen Netz regelmäßige Agent Advertisements verschickt und dem MN damit einen Zellenwechsel signalisiert. Das Heimatnetz besitzt als Komplement zum FA einen sogenannten Home Agent (HA). Er ist bei Abwesenheit des MN dafür verantwortlich, alle an den MN gerichteten Pakete abzufangen und an dessen Aufenthaltsort umzuleiten. Dafür bekommt der MN neben seiner statischen IP aus dem Heimnetz vom FA ähnlich zu DHCP eine neue, im fremden Netz gültige Adresse, die Care-Of-Address (COA). Sobald ein FA dem neuen MN eine COA zugeteilt hat, muss der HA darüber informiert werden. Innerhalb dieses Registrierungsprozesses ändert der HA den Eintrag in seiner Liste und sendet daraufhin alle Pakete, die den MN aus dem Heimatnetz erreichen sollen, an die entfernte COA (siehe Abbildung 4.1). In der Abbildung sind einige Router angedeutet, die alle vom einem fremden Kommunikationsgerät (CN - Correspondent Node) an den MN adressierten Pakete in das Heimatnetz leiten. Der HA sendet sie per Adressersetzung oder Tunneling in das fremde Netz weiter. Im neuen Netz übernimmt der FA die Zustellung an den MN. In diesem Bild ist das sogenannte Reverse Tunneling angedeutet, wo der HA auch als Bindeglied in der entgegengesetzten Richtung fungiert.

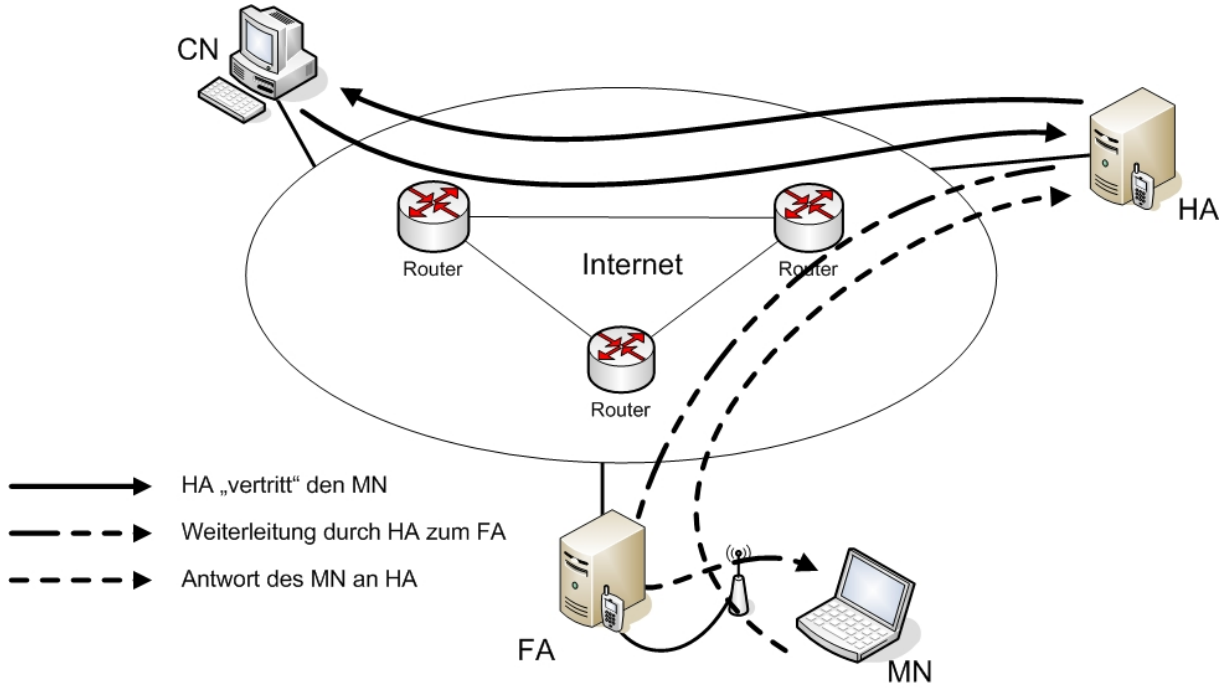

Abbildung 4.1: Mobile IP mit Reverse Tunneling

Weitere Möglichkeiten sollen hier verschwiegen werden, sind jedoch in [14] detailliert beschrieben. Offensichtlich wird der CN durch die Einführung des Mobile IP Protokolls nicht beeinflusst, für ihn bleibt das Protokoll transparent.

4.2.2 Probleme von Mobile IP

Das Konzept von Mobile IP ist immer dann vollkommen ausreichend, wenn die Frequenz von Zellenwechseln sehr niedrig ist. In diesem Zusammenhang kann man eher von einer drahtlosen Kommunikation sprechen, weil das Protokoll in quasi statischen Anwendungsfällen gut funktioniert. Im Fokus der Überlegungen zu Mikro-Mobilität stehen jedoch Situationen, in denen der MN häufige Zellenwechsel erfährt. Automatisch ist damit in Mobile IP jedoch ein aufwendiger, sich bei jedem Zellenwechsel wiederholender Registierungsprozess verbunden. Während eines solchen Zellenwechsels, im weiteren auch als Handoff bezeichnet, müssen Paketverluste und Verzögerungen in Kauf genommen werden. Gleichzeitig benötigen Anwendungen aber unter Umständen Zeitgarantien, z.B. bei Voice over IP (VoIP). Bei jedem Zellenwechsel treten zwei Arten der Verzögerung auf:

- Unter Move Detection Latency versteht man die Verzögerung t_{MD}, die durch das Erkennen eines Zellenwechsels auf der IP-Schicht entsteht. Da die Schichten im ISO/OSI-Referenzmodell strikt voneinander getrennt sind, ist es nicht beabsichtigt, daß die Schicht 3 von einem Wechsel des Point of Attachment informiert wird. Um also einen Zellenwechsel zu registrieren, sind in Mobile IP zwei Möglichkeiten vorgesehen: In regelmäßigen Zeitabständen t_{Adv} sendet ein FA sein Agent Advertisement. Weiterhin kann der MN auf das Ende der COA-Lebenszeit t_{COA} warten,

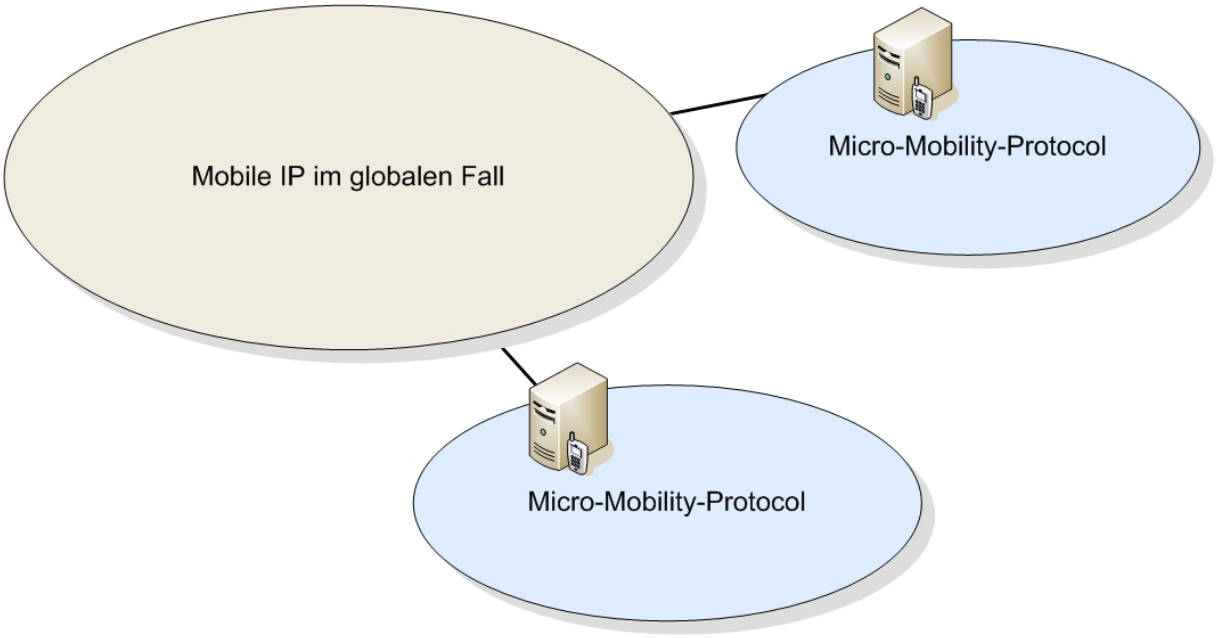

Abbildung 4.2: Die Trennung von Makro- und Mikromobilität

und dann ein neues Advertisement Request senden. Die Handoff-Erkennung auf der IP-Schicht verzögert sich um $t_{MD} := \min(t_{Adv}, t_{COA})$. Dabei stehen die Minimierung des Signal-Overheads innerhalb des Netzes im Widerspruch zur Minimierung der Move Detection Latency durch häufige Advertisements ($t_{Adv} \to 0$) oder geringe Lebenszeiten der COA ($t_{COA} \to 0$). Ansätze zur Lösung dieses Problems finden sich in 4.3.1.

- Die zweite Verzögerung, die Registration Latency [4] entsteht durch den Registrierungsprozess der COA beim HA. Dieser muß nach dem Konzept von Mobile IP bei jedem Zellenwechsel über die neue COA des MNs informiert werden. Da der HA im Allgemeinen nur über das Internet erreichbar ist, wird t_{Reg} sehr groß.

In dieser Zeit ($t_{MD} + t_{Reg}$) werden alle vom HA weiterzuleitenden Pakete an die alte, nunmehr falsche COA gesendet. Um wieder das Beispiel von VoIP zu nehmen, wären die Pakete auf dieser UDP-Verbindung verloren. Würde man TCP zugrunde legen, wäre das Problem sicher nicht behoben. Zum einen würde die zur Verfügung stehende Bandbreite des Netzes mehrfach beansprucht, zum anderen hätte man nun das Problem des Paketverlusts durch erhebliche Verzögerungen ersetzt. Um diesem Problem zu begegnen, sind diverse Ansätze veröffentlicht worden, die in den Punkten 4.3.2 bis 4.3.4 näher erläutert werden.

4.2.3 Definition von Micro-Mobility

Die im Konzept eines Mobile IP Handovers begründeten Verzögerungen waren der Initiator für die Abgrenzung der Begriffe der Mikro- und Makro-Mobilität und die daraus

resultierenden neuen Ansätze. Der sogenannte „Micro Mobility Approach“ [4] trennt die Begriffe wie folgt: Nach dem Schema regelt das Mobile IP Protokoll den Wechsel zwischen voneinander weit entfernten oder durch das Internet getrennten Netzen nach oben vorgestelltem Prinzip, man spricht in dieser Dimension von Makro-Mobilität. Die zugrundeliegende Netztopologie im makromobilen Bereich, bei der davon auszugehen ist, das regelmäßige Domänenwechsel eher die Ausnahme sind, macht Optimierungsversuche, auf denen im folgenden Wert gelegt wird, weitgehend unnötig, da das Verhältnis zwischen Aufwand und Nutzen gering wäre. Dementsprechend wird der Makro-Bereich in dieser Arbeit nicht mehr näher betrachtet. Im Gegensatz dazu werden einzelne Subnetze mit gemeinsamen Eigenschaften zusammengefasst (siehe Abbildung 4.2). Während sich der MN innerhalb dieser Netze bewegt, übernimmt ein Micro Mobility Protokoll die Lokalisierung des MNs. Mithilfe dieser Trennung können speziellere Lösungen gefunden werden, die bei Mobile IP aufgrund des Allgemeingültigkeit und Dimension nur ansatzweise realisierbar wären. Dieses Micro-Mobility Protokoll muss für Mobile IP im makro-mobilen Bereich transparent bleiben. Auf Grunglage dieser Annahme haben sich viele Protokolle entwickelt, die Probleme von Mobile IP mit Ansätzen wie Fast Handoff, Paging oder Hierarchical Mobility angehen.

4.3 Optimierungsansätze von Micro-Mobility

4.3.1 Schnelle Handoff-Erkennung

Unter Handoff-Optimierung versteht man Ansätze, die Zeit bis zum vollständigen Abschluss eines Zellenwechsels $t_{Handover}$ zu minimieren. Die IETF Mobile IP Working Group hat dazu einige Ansätze zusammengestellt:

Ziel all dieser Ansätze ist es, von der Schicht 2 einen Hinweis zu bekommen, damit das meist auf schicht 3 angesiedelte Micro-Mobility Protokoll schneller reagieren kann. Bei der sogenannten Layer 3 Movement Detection versucht man Informationen über den neuen FA zu erhalten, bevor ein Handoff auf Ebene 2 stattfindet. Dabei wird stets von der Lockerung der Grenzen zwischen den Schichten 2 und 3 gesprochen, um Handover auf Ebene 3 durch ein Handoff auf der darunterliegenden Ebene zu initiieren. Dieser Ansatz ist nicht trivial, zumal er die strikte Schichtentrennung des ISO/OSI-Referenzmodells übergeht, und damit zusätzlich die generelle Applikabilität reduziert. Andererseits können von Schicht 2 getriggerte Handover große Effizienzsteigerungen erzielen, z.B. in Signalstärkebasierten Handover-Schemata. Aufgrund der zahlreichen verschiedenen mobilen Geräte und Standards ist es jedoch kaum möglich, die Trennung zwischen den Schichten im allgemeinen Fall aufzubrechen, ohne dabei auf linkspezifische Eigenheiten eingehen zu müssen. Andrew T. Campbell [2] spricht in diesem Zusammenhang von dem Bedarf einer „Open Radio API“, um die Gemeinsamkeiten der verschiedenen Systeme herauszustellen und von linkspezifischen Details abstrahieren zu können. Ein elementarer Teil weiterer Lösungsansätze im Bereich des Handoffs sind Buffer- und Forwardingtechniken während des Handovers, wobei Pakete auf mehreren Wegen geroutet, teilweise „verlangsamt“ und somit an verschiedenen FAs empfangen werden können. Dadurch würde sich die Dauer des

Paketverlustes theoretisch (vgl. 4.4.3) auf die Länge des L2-Handoffs beschränken, also den Wechsel des Accesspoints.

4.3.2 Hierarchical Mobility

Hierarchical Mobility Management etabliert den bereits angesprochenen Ansatz in Micro-Mobility-Protokollen, lokale Bewegungen von MNs ausschließlich lokal zu behandeln. Dazu wird das Netz hierarchisch zusammengefasst und mit diversen Knotenpunkten unterschiedlicher Funktion versehen. Eine solche Struktur ist in Abbildung 4.3 dargestellt. Während sich der MN in diesem eng begrenzten Netz frei bewegt, kann er seinen WIP-POA wechseln, ohne daß ein Knoten ausserhalb dieses Netzes informiert werden müsste, was zur Folge hat, daß der Wechsel für den HA transparent bleibt. Dadurch entfällt die zeitaufwendige Registrierung beim Home Agent, die bei Bewegungen im Subnetz somit überflüssig ist, da fast die gesamte Routing-Strecke, nämlich bis zum Gateway, identisch geblieben ist.

Innerhalb des Netzes wird dann ein Micro-Mobility Protokoll die Zustellung der Pakete zum aktuellen Accesspoint übernehmen. Dazu benutzen die Protokolle eine sogenannte Location Database, in der die MNs auf Informationen zum Aufenthaltsort abgebildet werden. Die meisten Protokolle setzen dafür Knoten vorraus, die eine eigene Mobile-Routing-Tabelle führen, und diese für jedes ankommende Paket durchsuchen. Den Tabelleneinträgen werden Timer zugeordnet, sodaß sie regelmäßig aktualisiert werden müssen, oder nach einer vordefinierten Zeit t_{TTL} entfernt werden. In Micro-Mobility Protokollen unterscheidet man zwei grundlegende Konzepte, die unter dem Begriff „Hierarchical Mobility“ zusammmengefasst werden:

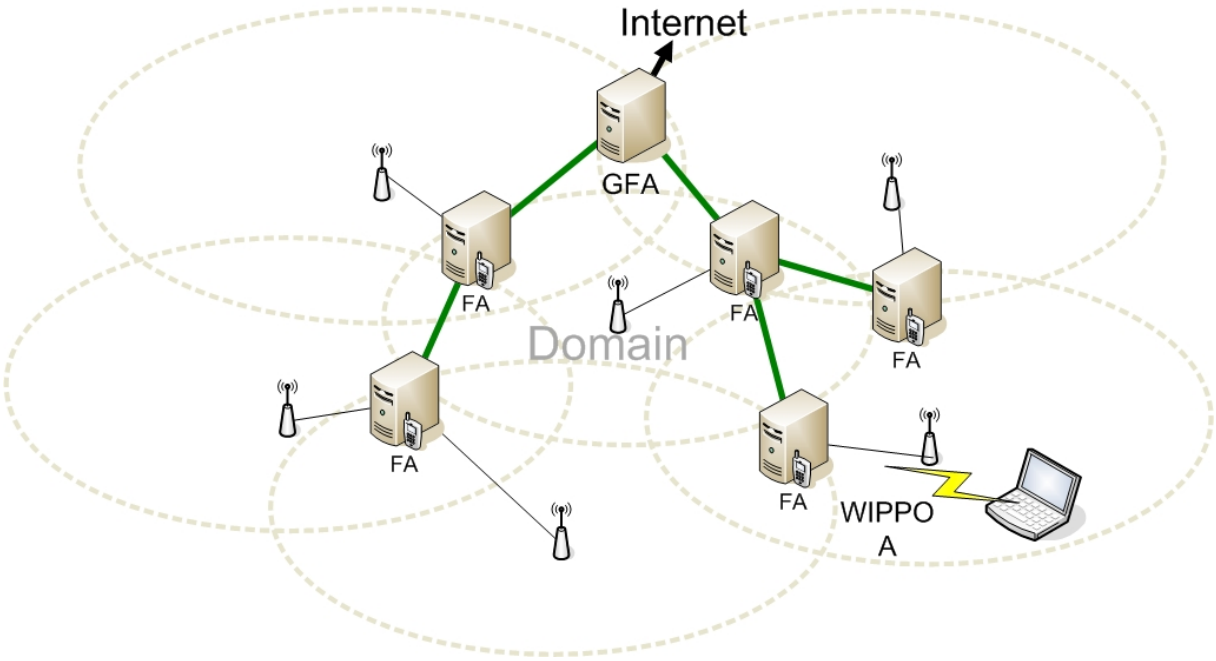

Abbildung 4.3: Eine Domäne mit Gateway und Foreign-Agent Hierarchie

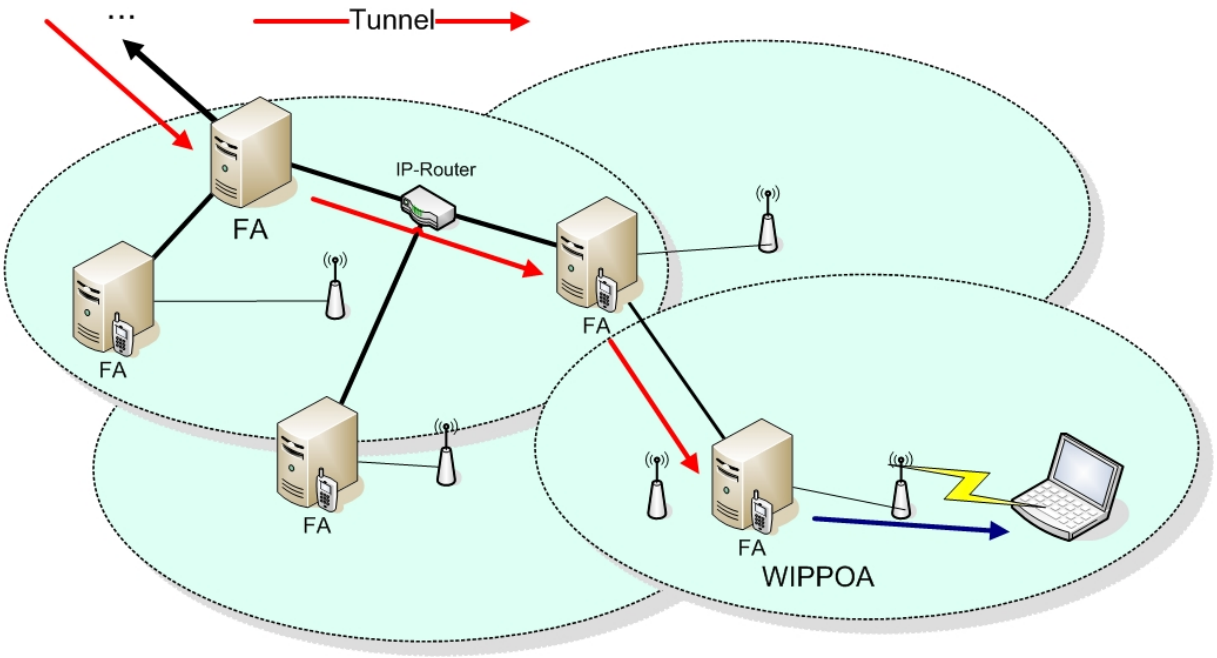

Abbildung 4.4: Das Prinzip von Hierarchical Tunneling

Hierarchical Tunneling

Wie der name schon sagt, ist dies ein Ansatz, nach dem alle relevanten Knoten in der Hierarchie mit einzelnen Tunneln verbunden sind, an dessen Enden der Gateway bzw. der MN sind. Die Datenbank ist auf einige Knoten innerhalb des Netzes verteilt. Zwischen den FAs werden die Pakete getunnelt, während sie am jeweiligen FA entkapselt und mit einer neuen Ziel-IP, also der des folgenden FAs wieder per IP-in-IP-Kapselung getunnelt werden. Dies hat den Vorteil, daß Protokolle nach diesem Konzept problemlos in bestehende Netze integriert werden können, da das normale IP-Routing die gekapselten Pakete an den FA weiterleitet (siehe Abbildung 4.4). Zur Aktualisierung der jeweiligen Einträge werden vom MN spezielle Registration Messages gesendet, damit die Einträge an den speziellen Knoten aktualisiert werden. Ein Protokoll, das auf diesem Prinzip basiert, ist Hierarchical Mobile IP (Kapitel 4.4.1).

Mobile-Specific Routing

Mobile-Specific Routing vermeidet den Aufwand, Pakete an jedem FA entkapseln und nach Suche des Folge-FAs wieder kapseln zu müssen. Statt dessen wird Routing benutzt, um die Pakete an den jeweiligen WIPPOA des MN zu leiten. Da diese Protokolle auf Tunneln verzichten, müssen alle Router dieser Systeme auch erweitertes Mobility-Routing beherrschen und sind daher nicht in bestehende Systeme integierbar. Um die Routingeinträge zu aktualisieren, werden zwei Ansätze unterschieden:

- implizite Aktualisierungen durch die Pakete vom MN

- explizite Signalisierung durch Registration Messages

Beispiele für Protokolle, die auf Mobile-Specific Routing basieren, sind Cellular IP und Hawaii, die im Kapitel 4.4 vorgestellt werden.

4.3.3 Paging

Während man bei fest mit dem Internet verbundenen Geräten wie einem Desktop-PC davon ausgeht, daß er stets erreichbar ist und dabei über praktisch beliebige Energiereserven verfügt, ist dies bei der Betrachtung von mobilen Geräten nicht der Fall. Insbesondere mobile Kommunikationsgeräte zeichnen sich dadurch aus, daß sie nur endliche, sogar recht geringe Reserven nutzen können und die meiste Zeit nicht aktiv kommunizieren. Paging betrachtet daher die Minimierung des Energieverbrauchs unter dem Aspekt der optimalen Netzanbindung. Denn auch der Nutzer von Mobilkommunikationsgeräten erwartet einen ähnlichen Service wie bei fest verkabelten Geräten, jederzeit quasi unmittelbar auf beliebige Internetressourcen zugreifen zu können [2, S. 46] und u. U. stets erreichbar zu sein. Man kann davon ausgehen, daß ein MN nur zu einem Bruchteil der Zeit, in der er eingeschaltet ist, tatsächlich kommuniziert. In der Zwischenzeit wird bei Mobile IP stets Energie verbraucht, um mindestens alle t_{TTL} Sekunden eine erneute Registierung im aktuellen Netz durchzuführen, damit es stets die aktuelle Position kennt.

Wenn man jetzt voraussetzt, daß es ausreicht, wenn das Netz nur noch einen nahezu beliebig ungenauen Standpunkt kennt, dann muss der MN sehr viel seltener kommunizieren. Also wird dem MN erlaubt, in einen Zustand zu wechseln, in dem er nicht gezwungen ist, regelmäßige „Registration Messages“ zu senden, sondern sich passiv zu verhalten. Ein möglicher Ansatz ist es, das zugrunde liegende Netz in Paging-Bereiche einzuteilen. Während sich der passive MN innerhalb dieser Paging-Area (PA) bewegt, ist eine erneute Registrierung nicht notwendig. er muß lediglich über die aktuelle PA informieren und auf eingehende Verbindungen warten. Wenn eine solche kommt, wird durch wieder eine aktive Bindung an das Netz hergestellt. In der Zwischenzeit genügt es, wenn der MN den Wechsel der PA bemerkt. Paging-Verfahren sind insbesondere in dem Protokollen Cellular IP und Hawaii (siehe Kapitel 4.4) beschrieben, aber auch für das Hierarchical Mobile IP Protokoll, das in 4.4.1 beschrieben ist, gibt es eine Paging-Erweiterung.

4.3.4 Fast Security

Um es vorweg zu nehmen: Man kann keine generelle Aussage über den Bedarf an Sicherheit in mobilen Netzen treffen. Sicherlich wird in fast allen denkbaren Szenarien unter Anwendung von mobilitätsorientierten Protokollen eine Form von Authentifizierung benötigt. Ebenso gibt es zahlreiche Anwendungen, in denen Daten über drahtlose Verbindungen wie auch in festverkabelten Netzen, unbedingt verschlüsselt werden müssen, in denen der Benutzer sich authorisieren muss, und Vorgänge zentral protokolliert werden müsssen. Gleichzeitig gibt es aber Anwendungsbereiche von mobiler Kommunikation, in denen das nicht erwünscht ist, zum Beispiel weil diese Mechanismen negativen Einfluss

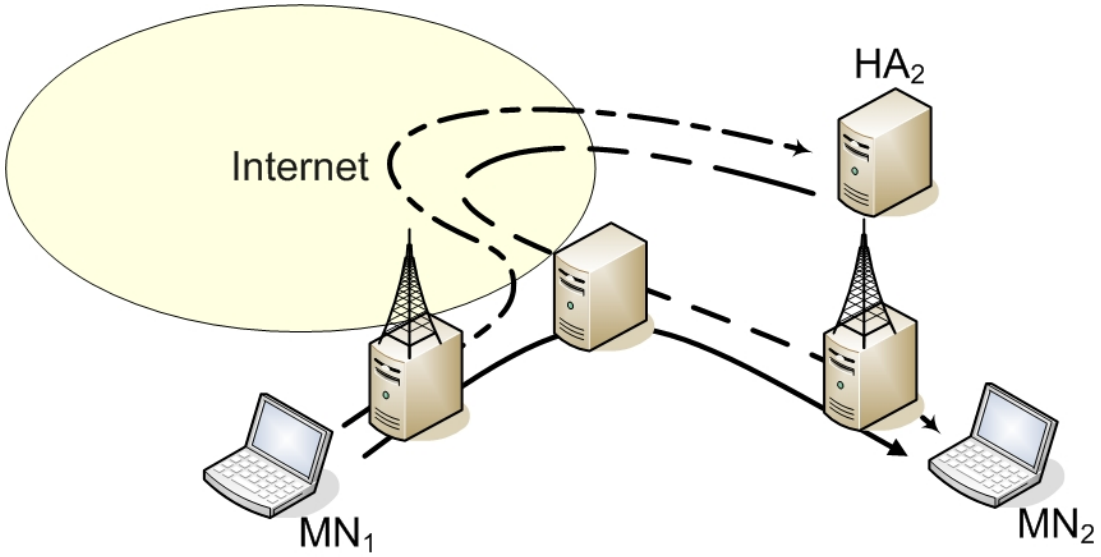

Abbildung 4.5: Das Triangular Routing Problem

auf Leistung, Quality of Service, Effizienz und Geschwindigkeit haben. Gleichzeitig hat die Möglichkeit, verschiedenste Aspekte von AAA innerhalb eines Micro-Mobility-Protokolls benutzen zu können, maßgeblichen Einfluss auf dessen praktische und allgemeine Anwendbarkeit. Gleichzeitig müssen bei all den Überlegungen auch Effizienz und Geschwindigkeit, Abhängigkeiten und Managementaufwand berücksichtigt werden. Zum Beispiel wären Abhängigkeiten von weiteren, möglicherweise weit entfernten AAA-Systemen [8] im Bezug auf Handoff-Geschwindigkeit kaum praktikabel, in dem Fall müssten schon zum Handoff-Zeitpunkt die nötigen Informationen am neuen Accesspoint verfügbar sein. Insbesondere durch regelmäßige Zellenwechsel wird deutlich mehr Kommunikation zwischen verschiedensten Stellen benötigt, was dem Ziel von Micro-Mobility-Protokollen zuwider läuft. Es müssen also Lösungen gefunden werden, die über das AAA-Konzept von Mobile IP hinausgehen. Ein Ansatz dafür ist das „Fast Session Key“ Verfahren [6], das im Kapitel 4.4.3 näher beschrieben wird.

4.3.5 Das Triangular Routing Problem

Im Zusammenhang mit Micro-Mobility beschreibt Triangular Routing einen Ansatz, unnötige Wege in der Kommunikation zu unterbinden. Für das Mobile IP Protokoll gibt es dazu eine Erweiterung zur Routenoptimierung. Wie in Abb. 4.5 kann es insbesondere in Netzen, in denen der größte Teil des Datenverkehrs innerhalb der Domäne bleibt, zu solchen ungünstigen Routen führen. In der Abbildung sei eine Kommunikation zwischen zwei MNs innerhalb eines Subnetzes dargestellt. MN_1 sendet an MN_2 über dessen HA, weil er ihn in seinem Heimatnetz adressiert. Es kann nicht davon ausgegangen werden, daß MN_1 mit einem erweiterten Mobilitätsprotokoll arbeitet, daher kann über einen Hinweis an MN_1 keine Optimierung stattfinden. Zahlreiche Micro-Mobility Protokolle haben sich dieses Problems angenommen, teilweise explizit in ihre Routingstrategie aufgenommen, meist jedoch implizit durch deren Konzept. Das Cellular IP Protokoll nutzt beispielsweise die Heimatadressen der MNs zur Adressierung innerhalb der Domäne. Das Triangular

Routing Problem wird damit implizit gelöst. Ein Beispiel für die explizite Behandlung ist das Fast Handoff Protokoll. Der betriebene Aufwand steht jedoch zum Nutzen nur dann in einem vernünftigen Verhältnis, wenn der Großteil der Pakete innerhalb der Domäne bleibt.

4.4 Die bekanntesten Protokolle

Es gibt eine Reihe verschiedener Konzeptstudien und Protokolle, die auf unterschiedlichen Gewichtungen der Ansätze basieren. Diese in ihrer Vielzahl vorzustellen, würde die Dimension dieser Arbeit weit übersteigen. Statt dessen werden hier nur einige ausgewählte Protokolle vorgestellt, die charakteristisch für Überlegungen bei Mikromobilität sind.

4.4.1 Hierarchical Mobile IP

Hierarchical Mobile IP [10] bildet eine natürliche Erweiterung zum Mobile IP Protokoll. Es wurde in Zusammenarbeit zwischen Nokia und Ericsson entwickelt und geht von einer Erweiterung des bekannten Mobile IP Modells aus, basierend auf Hierachical Tunneling. Dabei wird der dem HA bekannte FA von dem Gateway Foreign Agent (GFA) des jeweiligen Netzes repräsentiert (siehe Abbildung 4.3). Sobald der MN das Netz betritt, registriert er sich beim HA einmalig mit der IP-Adresse des GFA. Diese globale COA bleibt während des gesamten Aufenthalts in der Domäne konstant. Registration Messages von MNs erzeugen Tunnel zwischen jeweils benachbarten FAs nach dem Konzept von Hierarchical Tunneling. Dementsprechend kann der MN lokal über seine Heimatadresse identifiziert werden, da nur die Knoten mit Tunnelendpunkten involviert sind.

Zudem kommt ein Paging-Konzept in Hierarchical Mobile IP zur Anwendung, die in [11] als Erweiterung zu Mobile IP vorgestellt wurde, und nach folgendem Schema funktioniert:

Falls der MN gerade keine Daten sendet oder empfängt, kann er in den Idle-Modus wechseln. Wenn der MN inaktiv ist (im Folgenden IMN), ist die genaue Position in der aktuellen Domäne nicht mehr bekannt. Dafür werden Paging Foreign Agents (PFA) eingeführt, die jeweils ihre eigene Paging Area verwalten. Diese sind Unterbäume der FA-Hierarchie mit dem PFA als Wurzel. Jeder PFA führt eine Liste aller IMN in seiner Paging Area. Der PFA ist dann dafür zuständig, an den IMN adressierte Datenpakete zu verwalten und nach dem MN zu suchen. Damit das möglich, muss der IMN regelmäßig aktiv werden, um solche Anfragen oder bei einem Zellenwechsel auch andere Agent Advertisements zu bemerken. Macht er dies nach festgelegten Intervallen, dann spricht man von „Time Slot Based Paging“ [11]. Der Effekt ist zum einen, dass der MN durch Inaktivität Energie spart, zum anderen können durch dieses passive Verhalten Ressourcen des Netzes geschont werden, weil ein IMN das Netz nicht belegt. Zwar geht dies in Abhängigkeit vom Paging Intervall zu Lasten der Reaktionszeit, andererseits ist diese bei dem Aufbau einer Verbindung in mobilen Netzen im Gegensatz zur effektiven Energieverwaltung von untergeordneter Bedeutung.

4.4.2 Fast Handoff und Proactive Handoff

Basierend auf den Verbesserungen durch Hierarchical Mobile IP wurden im Fast Handoff Protokoll von Ericsson zwei weitere Probleme behandelt: Im Hinblick auf Realtime-Anwendungen wie VoIP sollte das Handoff-Management optimiert werden, zum anderen wollte man das Triangular Routing Problem (siehe 4.3.5) innerhalb der Domäne lösen.

Das Hierachical Mobile IP-Protokoll optimiert die Erkennung von Zellenwechseln nicht. Im Gegensatz dazu erwägt Fast Handoff die Möglichkeit der Interaktion mit der Radio-Schicht, um sich beim neuen FA registrieren zu können, bevor der Handoff überhaupt stattgefunden hat. Stets werden auf Schicht 2 Informationen über die umliegenden Netze gesammelt. Auf Grundlage dieser Daten ist es möglich, Zellenwechsel vorrauszusagen und das Protokoll darüber zu informieren. In [4] wird diese Interaktion mit dem Radio Interface als SHRT, Strong handoff Radio Trigger bezeichnet. Diese wird von allen beteiligten Knoten registriert und informiert die beteiligten über den MN, den alten und neuen FA. In Fast Handoff ist der MN somit in der Lage, sich über den alten FA bereits beim neuen FA zu registireren, wie in Bild 4.6 a) dargestellt. Die Registrierung läuft in vier Schritten ab:

1. MN empfängt Informationen über den bevorstehenden Zellenwechsel, informiert durch einen Trigger von Schicht 2 und sendet Request an FA_{neu}

2. FA_{neu} sendet ein Agent Advertisement über FA_{alt} zum MN

3. Der MN registriert sich bei FA_{neu} über den alten Radio Link. Mobile IP Bicasting kann zudem verwendet werden, um Paketverluste zu vermeiden (optional)

4. Wechsel des MN zum neuen WIPPOA

Hinter Proactive Handoff [13] steht ein ähnliches Prinzip, mit dem Unterschied, dass der IP Handoff nicht vom MN, sondern von den beiden FAs initiiert wird (vgl. Abbildung 4.6 b)).

1. FA_{neu} bekommt ein SHRT und sendet eine Registrierungsanfrage zum GFA, der daraufhin ein Bicasting zu FA_{alt} und FA_{neu} einleitet

2. FA_{neu} sendet ein Agent Advertisement über FA_{alt} zum MN

3. MN beginnt eine normale Registrierung beim neuen FA

Das Zeitfenster, in dem der MN nicht erreichbar ist, wird demnach auf die Dauer des Schicht 2 Handoffs beschränkt.

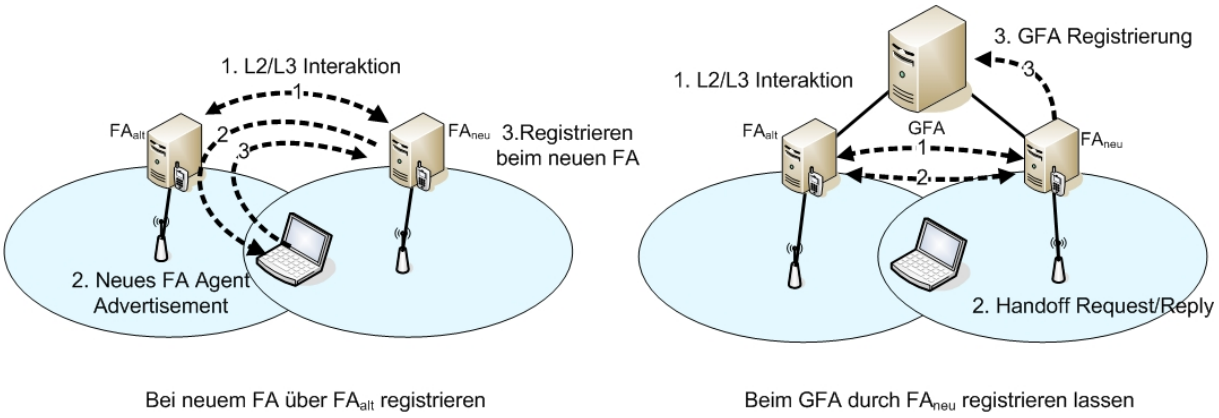

Abbildung 4.6: Handoff Mechanismen für a) Fast Handoff und a) Proactive Handoff [4].

4.4.3 Cellular IP

Das Cellular IP Protokoll hat eine besondere Stellung unter den Micro-Mobility Protokollen. Es verzichtet vollkommen auf Tunneling und Adresskonversion, indem es das bekannte IP-Routing innerhalb einer Domäne ersetzt. Es wird durch ein Netz sogenannter Cellular IP Router (CIPR) ersetzt, die auch fremde IPs im Netz korrekt routen können (All-IP Netzwerk). Ursprung dieses Ansatzes war folgende Überlegung: Man nehme ein einfaches Netz aus Accesspoints, Switches und einigen MNs. Die Lernfähigkeit von Ethernetswitches wird darin zur MN-Suche benutzt. Solch ein Konzept ist preiswert, einfach und effizient. Leider gibt es keine vernünftigen Möglichkeiten, um dieses Netz nach zuvor vorgestellten Ansätzen zu optimieren. Cellular IP nutzt aber nach diesem Konzept Datenpakete zum Aktualisieren von Locationtables, entweder auf Schicht 2 oder 3. Cellular IP Nodes übernehmen diese Verwaltung, das sind entweder einfache Switches (Schicht 2) oder Cellular IP Router (Schicht 3). Es wird behauptet, ein Cellular IP Netzwerk aus kostengünstigen Switches für mehrere tausend Benutzer noch effizient realisieren zu können [3, S. 179]. Das ist insbesondere dann interessant, wenn es sich um die Realisierung von Pico-Netzen handelt, beispielsweise um die drahtlose Vernetzung von Campus-Universitäten, in denen eine komplexe Hierarchie von Routern unnötig erscheint. Zu einem komplexeren CIP-Netz gehören folgende Komponenten: Die Basisstationen sind zum einen als Accesspoints der MNs zu sehen, gleichzeitig sind es meist selbst Cellular IP-Router. Alle Basisstationen einer Domäne sind mit dem Internet über einen gemeinsamen Gateway verbunden, der als eiziger FA-Funktionen besitzen muß. Während sich der MN innerhalb des Netzes befindet, nutzt er die IP des Gateways als seine Care-Of-Address. Ausserhalb des Gateways gilt wieder das Prinzip von Mobile IP. Die am Gateway ankommenden Pakete werden entkapselt und im Cellular IP Netz versandt, wobei der MN über seine Heimatadresse identifiziert wird. Für die korrekte Zustellung der Pakete sind die hierarchisch angeordneten CIP-Router zuständig. Die Aktualisierung der „Location Tables“ erfolgt durch Snooping normaler IP-Pakete, solange der MN sich nicht bewegt. Besondere Registration Messages werden nur dann gebraucht, wenn der MN sich zum ersten Mal registriert oder seinen WIPPOA ändert. Allerdings ist es theoretisch denkbar, daß der MN längere Zeit nicht sendet, obwohl er de facto erreichbar ist. Dies wäre z.B. bei einem Videostream per UDP der Fall. Daher werden auch in diesem Protokoll regelmässige Registration Messages gesendet, wobei de-

ren Intervall konsequenterweise kleiner sein muß als die Lebenszeit der Routingeinträge. Zusätzlich werden vom Gateway regelmässige Beacons innerhalb des Netzes gebroadcastet. Jeder CIP Knoten, der diesen empfängt, sendet ihn auf allen anderen Schnittstellen weiter, abschließend kennt jeder Knoten den Uplink, nämlich das Interface zum Gateway. Durch dieses Verfahren wird theoretisch eine beliebige Erweiterbarkeit des Netzes möglich, die nahezu ohne Administrationsaufwand auskommt. Neben diesem Konzept wird in Cellular IP klassisches Paging benutzt. Wenn der MN inaktiv ist, muß er bei einem Wechsel der Paging Area ein Paging Packet senden. Dieses wird zum Gateway geroutet und muß irgendwann den Knoten passieren, der den Paging Cache für diesen MN besitzt, damit dieser aktualisert werden kann. Das hat zur Folge, daß das Netz eine Baumstruktur besitzen muss.

Handoff

In Cellular IP werden zwei Arten von Handoff unterschieden, mit jeweils unterschiedlicher Gewichtung von Paketverlust, Verzögerung, Aufwand und Netzbelastung: Beide Lösungen berücksichtigen die Möglichkeit, per SHRT einen L2-getriggerten Handoff beginnen zu können.

Hard Handoff verzichtet auf aufwendige Kommunikation zwischen den Beteiligten und nimmt Paketverluste in Kauf. Während der MN mit einer Basisstation verbunden ist, sucht er nach weiteren mit besserer Signalstärke und initiiert gegebenenfalls einen Handoff. Als erstes wird also der Accesspoint gewechselt und ein „Route Update Packet“ in Richtung Gateway gesandt. Geht man zur Vereinfachung davon aus, daß die Zeit für den Wechsel des Accesspoint $t_{Handoff}$ zu vernachlässigen ist, dann beschränkt sich die Dauer des Paketverlustes auf $t_{Handover} = t_{CR,neu} + t_{CR,alt}$, wie in der Abbildung 4.7 dargestellt: Alle Pakete gehen verloren, die den Crossover Node (CR) vom Zeitpunkt des Handoffs bis zur Ankunft des „Route Update Packets“ durchlaufen haben, und alle, die sich zwischen CR und MN befanden. Der CR ist dabei der Knoten, an dem sich die alte und neue Route kreuzen, also im schlechtesten Fall der Gateway. Sicherlich ist diese Zeitspanne $t_{Handover}$ viel kleiner als t_{HA}, die Zeit, die im Abschnitt 4.2.2 als Verzögerung durch die Registrierung beim Home Agent für Mobile IP definiert wurde. Dadurch wird die Anzahl der Paketverluste bereits in diesem Ansatz stark reduziert.

Semisoft Handoff verfolgt das Ziel, Paketverluste vollständig zu vermeiden. Mithilfe sogenannter „Semisoft Packets“ werden die Routingeinträge zur neuen Basisstation bereits vor dem eigentlichen Handoff erzeugt. Dazu wird ein solches „Semisoft Packet“ über den neuen Accesspoint gesendet, wobei der MN danach sofort wieder auf die alte Basisstation zurückschaltet. Aufgrund des besonderen Pakets wird der CR daraufhin alle ankommenden Pakete an die alte und neue Base Station senden (Bicasting). Nach einem „Semisoft Delay“ wird dann ein normaler Handoff durchgeführt und das Bicasting aufgehoben. Damit sind Paketverluste jedoch noch nicht behoben. Es kann durchaus passieren, daß $t_{CR,alt} > t_{CR,neu}$ ist. Dann kommen die Pakete am neuen Accesspoint an , bevor der MN sie am alten empfangen kann. Selbst ein zeitloser Handoff ($t_{Handoff} \to 0$) wäre nicht schnell genug, um diese Pakete am

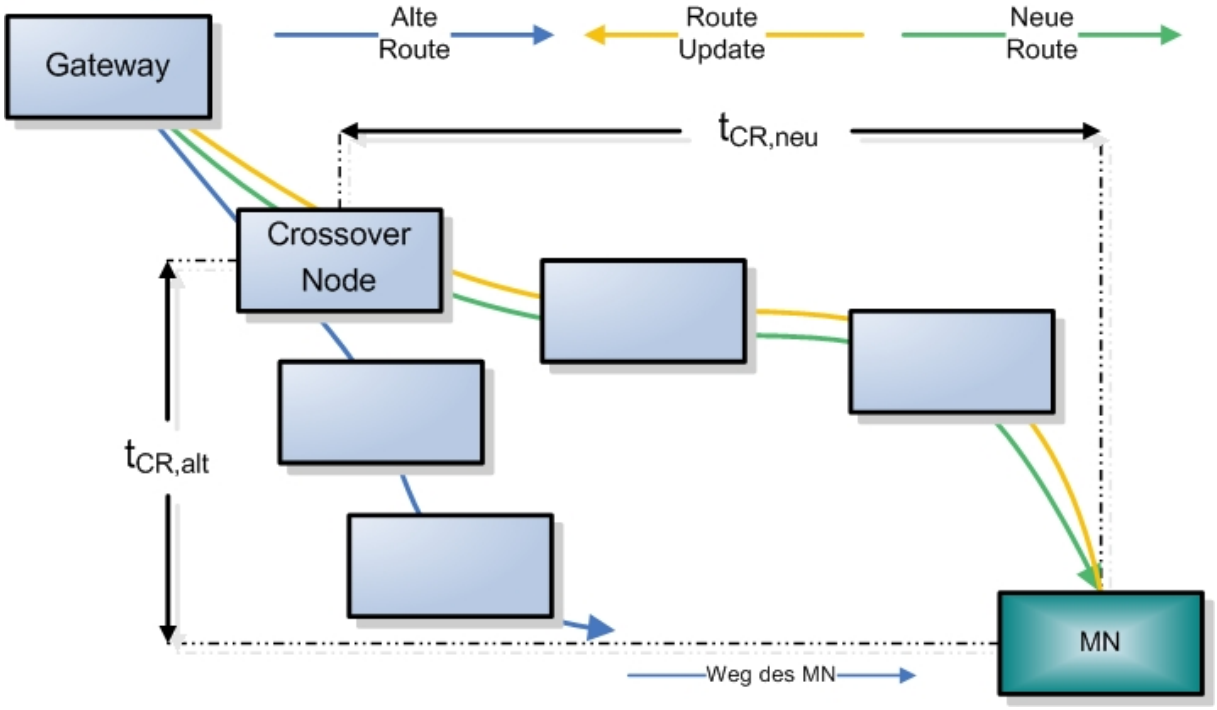

Abbildung 4.7: Ein Handoff mit Cellular IP über den Crossover-Node

neuen AP noch zu empfangen. Dieses Problem wird gelöst, indem die Pakete auf der neuen Route gebremst werden. Dazu wird ein konstanter Buffer für n_{CR} Pakete eingerichtet, der die Pakete bis zum Abschluss des Handoffs verzögert. Der Wert für n_{CR} bestimmt sich dabei einerseits aus bekannten Werten für das jeweilige Netz, z.B. anhand der explizit zu bestimmenden, maximal möglichen Länge eines Weges auf dem Baum der CIP-Hierarchie ausgehend vom jeweiligen CR, andererseits aus den Prioritäten zwischen Paketverlust und Delay, da n_{CR} bei jedem Handoff am CR konstant bleibt und somit stets auftritt.

Session Keys

Auch in Cellular IP is es von elementarer Bedeutung, die eintreffenden Location Update Pakete von jedem MN zu authentifizieren, damit nicht fälschlicherweise fremde Quellen die CIP-Routing-Informationen ändern können. Gefälschte Pakete können auf einfachste Art und Weise Angriffe auf das Netz darstellen. Daher dürfen nur authentifizierte Pakete Routinginformationen löschen, während normale Pakete bereits bestehende ohne weiteres aktualisieren dürfen. Mithilfe eines Session Keys wird das Problem der Authentifizierung sicher und effizient gelöst. Ein solcher Schlüssel berechnet sich bei CIP folgendermaßen:

- Die IP-Adresse IP_{MN}, also die Heimatadresse des MN
- Eine Zufallszahl R_{MN}, die dem jeweiligen MN bei der ersten Registrierung innerhalb des Netzes zugeteilt wird

- Einen privaten Schlüssel K_{CIP}, der nur den Basisstationen des Cellular IP Netzwerkes bekannt ist

Daraus berechnet sich der Session Key mit Hilfe einer MD5 Hashfunktion zu $Key_{session} := MD5(IP_{MN}, R_{MN}, K_{CIP})$. Dieser Session-Key wird dem MN zugeordnet, sobald er sich dem Netzwerk anschließt. Alle „Controll-Pakets“, die zum Erstellen einer neuen Route zum MN benutzt werden müssen, beinhalten den Zufallsschlüssel R_{MN} und einen „Timestamp“, um Replay zu verhindern. Im Payload des jeweiligen Pakets ist ein „Message Authentication Code“ enthalten, der sich aus dem Session Key, einem Zeitstempel und dem Paketinhalt berechnet. Somit können die Stationen die Echtheit der eintreffenden Pakete prüfen, ohne mit aussenstehenden Instanzen kommunizieren zu müssen, indem sie selbst den entsprechenden Session Key aus den im jeweiligen Paket enthaltenen R_{MN} und der IP-Adresse mit K_{CIP}) berechnen und daraufhin den Inhalt der Nachricht verifizieren können. Zusätzliche Sicherheit kann gewonnen werden, wenn der private Netzwerkschlüssel K_{CIP} innerhalb des Netzes regelmäßig geändert wird.

4.4.4 Hawaii

Das Handoff-Aware Wireless Access Internet Infrastructure Protokoll, kurz Hawaii, wurde von Lucent Technologies entwickelt. Es agiert als Micro-Mobility Protokoll in einer geschlossenen Domäne. Zwar ist es in seinen Fähigkeiten und Optimierungsansätzen ähnlich zu Cellular IP, dennoch gibt es einige konzeptionelle Unterschiede. Auch Hawaii verzichtet auf Tunneln innerhalb der Domäne, allerdings wird dem MN beim Betreten eine COA per DHCP zugewiesen. Diese bleibt konstant, solange sich der MN innerhalb dieses Netzes bewegt. Mit der „Path Setup Powerup Message“, also der ersten Registrierung innerhalb des Netzes, wird eine Nachricht zum „Domain Root Router“(DRR), gesendet. Dieser stellt die gemeinsame Schnittstelle des Netzes dar und ist der Tunnelendpunkt für Mobile IP. Alle Router auf dem Weg zum DRR nehmen die COA des MN in ihre Routing-Tabelle auf. Sobald der DRR die Nachricht bekommt, wird ein Acknowledge auf dem soeben erstellten Weg zum MN zurückgesandt und die Registrierung beim HA durchgeführt. Alle anderen Router innerhalb der Domäne kennen die IP des MN nicht und werden daher alle für den MN bestimmten Pakete, z.B. von anderen MNs in Richtung DRR senden. Path Setup Refresh Messages sorgen für regelmässige Aktualisierungen der Pfade, während Update Messages bei jedem Handoff benutzt werden, um neue Routingeinträge zu erstellen.

Hawaii bietet verschiedene Strategien, sogenannte „Path Setup Schemes“ an, mit unterschiedlichen Gewichtungen, Paketverluste und Verzögerungen zu minimieren und den Administrationsaufwand gering halten zu wollen. Diese unterteilen sich grundsätlich in zwei Gruppen:

- **Forwarding Path Setup Schemes** leiten die Pakete während des Handovers von der alten zur neuen Basisstation weiter, bevor der Crossover Router seinen Eintrag aktualisiert und somit direkt an den MN routet. Hawaii unterscheidet dabei zwei Ansätze:

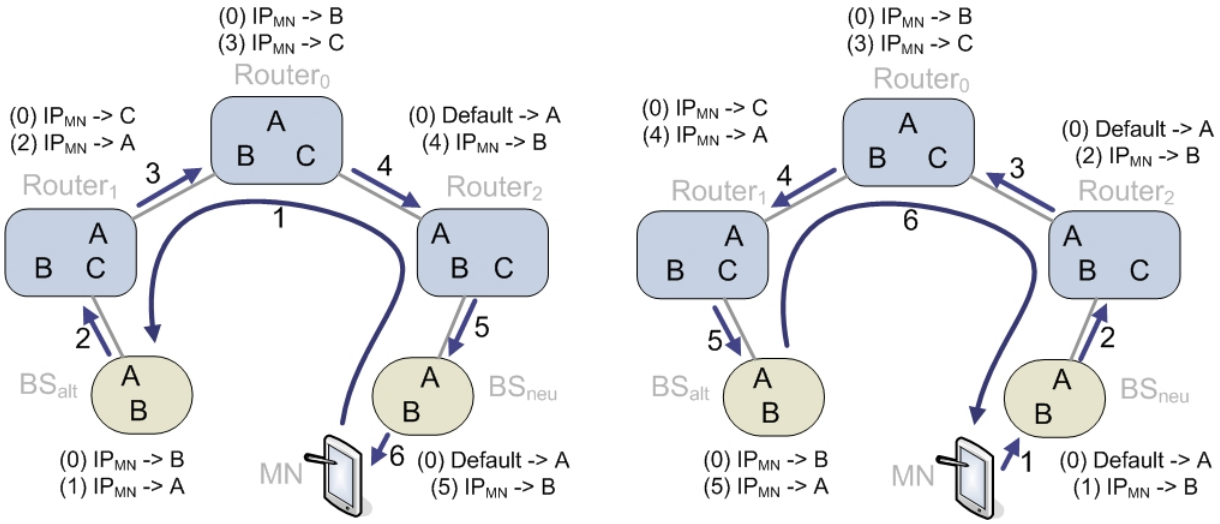

Abbildung 4.8: Forwarding Path Setup Schemata MSF (links) und SSF (rechts), aus [7]

Multiple Stream Forwarding (MSF) funktioniert folgendermaßen: Nach dem Handoff sendet der MN eine „Path Setup Update Message“ an die alte Basisstation, mit der IP der neuen. Sobald BS_{alt} das Paket empfangen hat, wird das Interface gesucht, unter dem es BS_{neu} erreichen kann und ändert den Routingeintrag für den MN entsprechend (siehe Abbildung 4.8 a)). Er sendet die Updatenachricht nun auf dem Pfad quasi zurück, nämlich zum $Router_1$. Dieser verfährt ebenso und sendet die Nachricht weiter. Sobald die Nachricht BS_{neu} erreicht, bekommt der MN eine Bestätigung. Es ist nicht unbedingt offensichtlich, daß bei diesem Verfahren mehrere Datenströme entstehen, die zu falscher Reihenfolge der Pakete führen können und dem Schema seinen Namen gegeben hat. Die Entstehung soll anhand der Verbindung zwischen $Router_1$ und BS_{alt} erläutert werden. Nachdem BS_{alt} das Paket 2 abgeschickt hat, passieren noch einige Pakete $Router_1$. Durch Paket 2 ändert $Router_1$ seine Routingtabelle, und alle ankommenden Pakete werden zu Interface A geleitet. Erst danach kommen jedoch falsch geroutete Pakete an Interface C an, die anschließend - in der falschen Reihenfolge - über A weitergeleitet werden. Da dieser Effekt an jedem Knoten bis zum CR auftritt, kann dies negative Folgen für VoIP oder auch TCP-Verbindungen haben. Paketverluste werden so jedoch vollständig vermieden.

Single Stream Forwarding (SSF) soll dieses Problem beheben. Dazu führt einen neuen Begriff, das „Interface Based Forwarding“ ein, um die korrekte Paketreihenfolge der umgeleiteten Pakete einzuhalten, ohne dafür Tunnel nutzen zu müssen. Eine normale Routingfunktion bildet eine IP-Adresse auf ein Interface ab ($IP \rightarrow Interface_{out}$). Die hier zum tragen kommende Erweiterung des Routingverfahrens entspricht der Funktion $Interfaces_{in} \times IP \rightarrow Interface_{out}$ und ähnelt in ihrem Effekt der ergänzten Routenoptimierung bei Mobile IP. Mit dieser Erweiterung kann ein einzelner Stream der von BS_{alt} zu BS_{neu} weitergeleiteten Pakete realisiert und Paketverluste verhindert werden. Allerdings ist der Aufwand für diese Alternative relativ hoch, sodaß sie sich bei den wenigsten Handoffszenarien lohnt.

- **Nonforwarding Path Setup Schemes** leiten die Pakete bereits am Crossover-Router um, ohne sich selbst um die Vermeidung von Paketverlusten durch Forwarding zu kümmern. Diese Paketverluste werden entweder in Kauf genommen, oder können durch parallele Kommunikation mit BS_{alt} und BS_{neu} abgefangen werden. Letzteres wäre durchaus denkbar für CDMA-basierte Kommunikationssysteme oder WaveLAN [7]. Zwei Schemata realisieren diesen Ansatz:

  Das Unicast Nonforwarding Schema (UNF) setzt voraus, daß ein MN kurzzeitig mit der alten und neuen Basisstation kommunizieren kann. Dazu sendet der MN eine Update Message an die neue Basisstation, die daraufhin einen Eintrag für den MN erzeugt. BS_{neu} sendet diese Nachricht weiter zum nächsten Router auf dem Weg zu BS_{alt}. Falls letztere das Paket empfängt, wird direkt ein Acknowledge an den MN gesandt, da alle bis dato vorhandenen Routingeinträge für den MN geändert wurden. Offensichtlich findet hier das Update der Routingeinträge bereits auf dem „Hinweg“ statt, sodaß die Pakete schneller auf dem neuen Weg geroutet werden. Dementsprechend ist ein Forwarding noch ankommender Pakete nicht möglich, weil das Route Update Paket erst nach diesen an der alten BS ankommt. Durch das fehlende Forwarding ist dieses Konzept unter vorgenannten Bedingungen optimal.

  Das Multicast Nonforwarding Schema ist prinzipiell identisch zu UNF, mit dem Unterschied, daß der CR Pakete kurzzeitig multicastet, um Paketverluste in Fällen zu verhindern, in denen der MN nur jeweils von einer BS emfangen kann. Dieses Schema ist entsprechend für „Time Division Multiple Access“ Netzwerke gedacht.

Transparenz des Hawaii-Protokolls

Dass die Micro-Mobility-Protokolle gegenüber dem Mobile IP Protokoll nach außen transparent sein müssen, wurde bereits vorab herausgearbeitet. Das Hawaii-Protokoll ermöglicht es u.a. sogar, daß Hawaii auch für den MN transparent bleiben kann. Um dies zu realisieren, muss die jeweilige Basisstation Mobile IP Nachrichten in Hawaii-konforme Nachrichten übersetzen. Der MN sende eine Mobile IP Registration Message, um sich beim HA zu registrieren. Die entsprechende Basisstation übersetzt dieses Paket in eine Path Setup Powerup Message, die einen oben beschriebenen Registrierungsprozess auslöst. Desweiteren wird der MN regelmässige Advertisement Request absenden, um nach neuen FAs zu suchen. Aus diesen erzeuge die Basistation eine „Path Setup Refresh Message“, falls ihr der MN bereits bekannt ist. Damit werden die Routingeinträge innerhalb der Domäne aktualisiert. Falls die Basisstation den MN nicht kennt, oder eine Nachricht auf einem anderen Interface erwartet hätte, muß eine „Path Setup Update Message“ gesendet werden, um neue Routingeinträge zu erstellen. Der MN bekommt davon jedoch nichts mit.

	Ethernet Switch	**Cellular IP**	**Hawaii**	**Hierarchical MIP**
MRP Schicht	L2	L3	L3	„L3.5“
MRP	Alle Switches	CIP Knoten	alle Router	alle FAs
MN ID	MAC	Home IP	COA	Home IP
Ohne Wirkung	-	Switches	Switches	Router
MRP „ID“	(L1)	MAC	MAC	IP
Updates	implizit	(implizit)	explizit	explizit

Tabelle 4.1: Ein Vergleich bekannter Micro Mobility Protokolle

4.5 Vergleich der Protokolle

4.5.1 Konzepte

Ganz offensichtlich gibt es viele Ähnlichkeiten in der Funktionsweise von Micro Mobility Protokollen. Im Detail ergeben sich jedoch zahlreiche charakteristische Unterschiede wie die Identifizierung der MNs, die Struktur und Verteilung der „Location Databases“, sowie die Möglichkeiten der Datenbankaktualisierung [8]. Solche Einträge sind mit Timern versehen, und müssen durch regelmässige Nachrichten aktualisiert werden. Das kann implizit durch das Snooping normaler Pakete geschehen, oder explizit durch besondere Nachrichten. Für ankommende Pakete bildet diese Serie von „next-hop“ Einträgen den Weg zum aktuellen WIPPOA. Dafür wird an Knoten des Netzwerks die Zieladresse ausgelesen, mit der eigenen Liste verglichen und zum nächstzen Knoten weitergeleitet. Diese Knoten werden im Weiteren als MRP, Mobile Routing Points bezeichnet. Im Falle von Hierarchical Mobile IP gibt es nur einige explizite MRPs, nämlich die FAs. Diese sind durch Tunnel miteinander verbunden und entkapseln die Pakete nur kurzzeitig, um als Router zu fungieren, indem sie den Adressaten auslesen und die Pakete daraufhin weiter tunneln. HMIP wird daher als ein „Schicht 3.5“ Protokoll bezeichnet. Auch Router zwischen den FAs beeinflussen die Weiterleitung der Pakete nicht (siehe auch Tabelle 4.1). Mit dieser IP-in-IP Kapselung enthält das Paket stets die IP des MN und als Absender und Empfänger die IPs der beiden beteiligten FAs. Bei Hawaii und CIP wird kein Tunneln benutzt. Statt dessen ist diese Information in den Paketen implizit enthalten, wobei Cellular IP als Identifikation der MNs die Heimatadresse benutzt. Dies führt dazu, daß zwischen den CIP Routern keine „normalen“ L3-Knoten existieren dürfen, da diese mit der IP des MNs nichts anfangen können und das Paket zurück in Richtung Gateway senden würden. In der Tabelle 4.1 werden daher nur jene Knoten als „Ohne Wirkung“ bezeichnet, die keinen Einfluss auf die Zustellung der Pakete haben. Alle soeben genannten Merkmale hängen unmittelbar von der Entscheidung ab, auf welcher Schicht das Protokoll agieren soll. Ein weiterer in der Tabelle aufgeführter Unterschied ist die Realisierung der Routing Updates. Diese Entscheidung ist von den vorhergehenden vollkommen unabhängig. Cellular IP versucht hier, eine implizite Signalisierung durchzusetzen, um ohne Verwaltungsnachrichten auszukommen. Um die Protokolle zu strukturieren, können sie auch wie in der Tabelle 4.2 charakterisiert werden.

OSI-Schicht	Update durch	
	Pakete	**Nachrichten**
2	Ethernet Switch	-
3	Cellular IP	Hawaii
3.5	-	HMIP

Tabelle 4.2: Matrix zur Charakterisierung von Protokollen

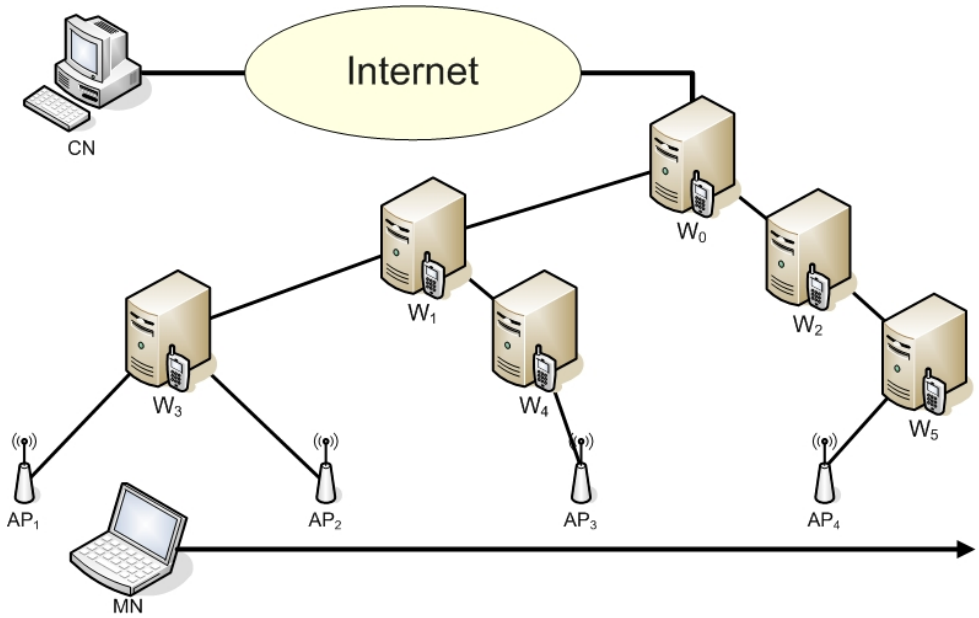

Abbildung 4.9: Das Netzwerk der Simulation

4.5.2 Simulation

Es ist nicht einfach, Protokolle zu vergleichen, die auf so unterschiedlichen Annahmen und Vorraussetzungen basieren wie das Hierarchical Mobile IP Protokoll, Hawaii und Cellular IP. Verschiedenste Quellen haben in Versuchsaufbauten die Leistungsfähigkeit der einzelnen Protokolle herauszustellen versucht. In einer Veröffentlichung der IEEE [7] stellt der Author Dr. R. Ramjee das Hawaii Protokoll vor, um es in einer Simulation anschließend mit dem Mobile IP Protokoll zu vergleichen und die Unterschiede von MSF, SSF, UNS und MNF herauszustellen. Andrew T. Campbell, Professor an der Columbia University, New York, richtete seinen Artikel zum Thema Internet Micromobility [6] auf das Cellular IP Protokoll aus, dessen Entwicklung er maßgeblich vorantrieb. Dabei legte er in seinem Versuchsaufbau eine etwas andere Struktur des Netzwerkes fest als Dr. R. Ramjee. Daß die Festlegung des zugrundeliegende Versuchsaufbaus nicht trivial ist, liegt nahe. Um in dieser Arbeit die drei oben genannten Protokolle vergleichen zu können, muß ein Aufbau gefunden werden, der allen Protokollen gleichmässig gerecht werden kann. Das ist nicht ohne weiteres möglich. Daher werden in dieser Arbeit die Ergebnisse weniger, dafür möglichst unterschiedlicher Simulationen aus dem Artikel „Comparison of IP Micro-Mobility Protocols“ [8] vorgestellt und mit eigenen Überlegungen erweitert.

	CIP	Hawaii	HMIP
W_0	Gateway	DRR	GFA
W_i	CIP Nodes	Hawaii Router	Router
AP_i	CIP Nodes	Hawaii Router	FA

Tabelle 4.3: Die Knoten der einzelnen Simulationen

Paketverluste durch Handoff

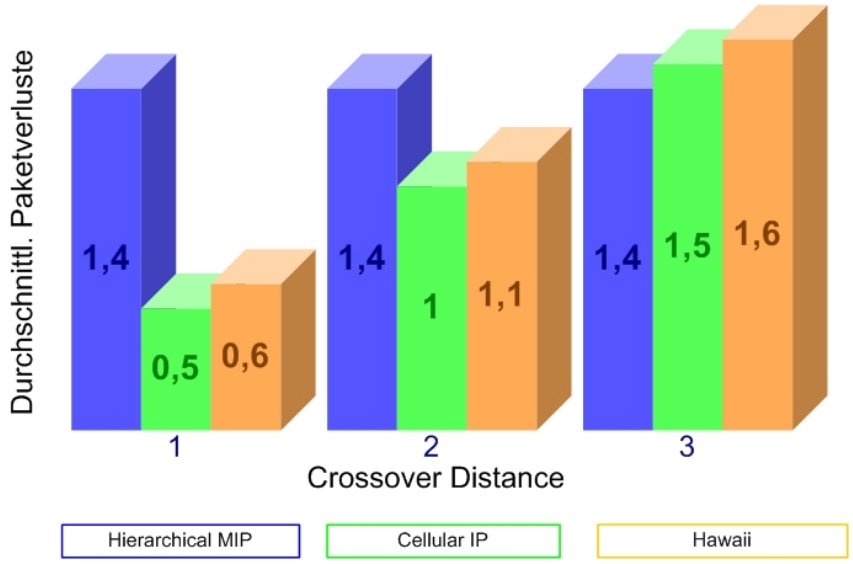

Abbildung 4.10: UDP Paketverluste während des Handoffs, entnommen aus [8]

Im ersten Versuch sollen die drei Protokolle im Bezug auf Paketverluste während des Handoffs untersucht werden, basierend auf der Netztopologie aus Abbildung 4.9. Verglichen werden CIP Hard Handoff und Hawaii UNF mit Hierarchical Mobile IP basierend auf einem UDP-Datenstrom von 100 Packets/s zwischen CN und MN. Dabei wurde die Bewegung des MN im Versuchsaufbau so simuliert, das jeweils eine, zwei und drei Sprünge zwischen den APs und dem CN waren, also Wechsel zwischen $AP_1 \rightarrow AP_2$, $AP_2 \rightarrow AP_3$ und $AP_3 \rightarrow AP_4$. Dabei wurden die durchschnittlichen Ergebnisse nach 100 Versuchen im Diagramm 4.10 dargestellt. Es fällt auf, daß die Verluste für HMIP unverändert hoch bleiben, während sie bei Cellular IP und Hawaii jeweils sehr ähnlich sind, aber mit geringerer Entfernung zum Crossover Node sinken. Das zweite Ergebnis überrascht nicht, da die Handoffverzögerung bei CIP und Hawaii maßgeblich vom Anstand zum CR abhängt. Sobald der Routingeintrag in diesem Knoten aktualisiert wurde, werden die Pakete in beiden Protokollen wieder korrekt weitergeleitet. Anders verhält es sich bei HMIP, weil die Routingeinträge erst dann korrigiert werden, wenn die neue Registration Message den GFA erreicht hat. Der Abstand zum GFA ist in der Simulation jedoch konstant, und somit auch die duchschnittlich verlorenen Pakete. HMIP hat in diesem Zusammenhang also ein konzeptionelles Defizit, denn es kann von möglicherweise geringem Abstand zum CR nicht profitieren.
Es wurde festgestellt, daß CIP und Hawaii sich in dem Versuchsaufbau sehr ähnlich verhalten haben. Diese Ähnlichkeit geht verloren, wenn nicht mehr von einer Baumstruktur

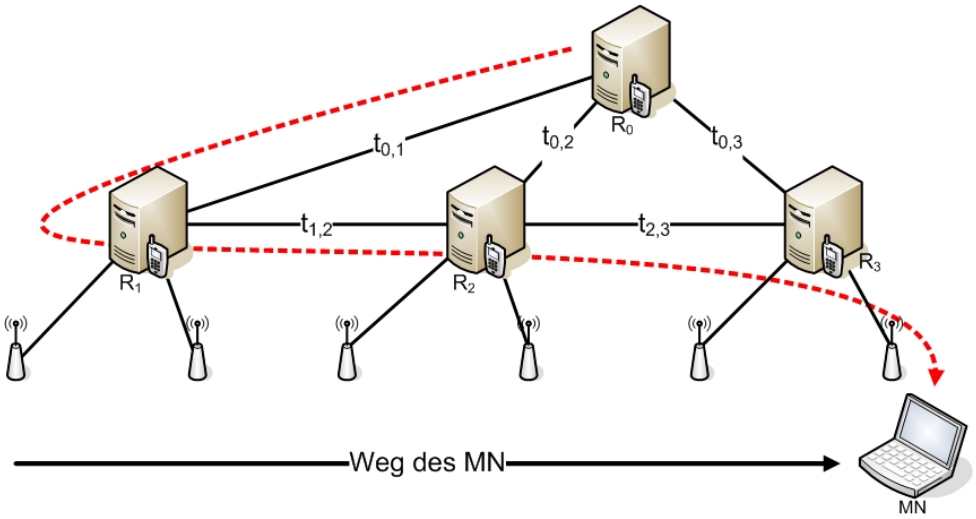

Abbildung 4.11: Eine suboptimale Route nach mehreren Handoffs

des Netzes ausgegangen wird, wie es zur Vereinfachung im Versuchsaufbau (Abb. 4.9) angenommen wurde. Wenn das nicht der Fall ist, macht sich bemerkbar, daß sich die Definition des Crossover Nodes zwischen CIP und Hawaii unterscheiden: Bei CIP ist der CR der erste gemeinsame Knoten der Wege vom alten und neuen CIP Node zum Gateway. Im Falle von Hawaii jedoch bestimmt sich der CR aus dem ersten gemeinsamen Knoten zwischen dem Weg vom alten AP zum DRR und dem kürzesten Weg zwischen den beiden Accesspoints. Das Ergebnis ist im Bild 4.11 dargestellt und zeigt eine schlechte Route nach fünf Handoffs unter der Annahmne, daß $t_{i,j} \leq t_{0,i} + t_{0,j}$ ist. Erst mit der nächsten „Path Refresh Message“ könnte die Route korrigiert werden, da der jeweilige WIPPOA die „Path Setup Refresh Messages“ auf dem kürzesten Weg zum DRR sendet. Falls die eingezeichnete Route gegenüber dem direkten Weg zwischen den Routern R_3 und R_0 (aus)suboptimal ist, gilt $t_{CR,3} \leq t_{0,1} + t_{0,2} + t_{0,3}$. Das bedeutet gleichzeitig, daß zusätzlich zu den während des Handoffs auftretenden Verzögerungen und/oder Paketverlusten bei der Routenkorrektur nachträglich auch die Reihenfolge der Pakete durcheinander gerät.

Webbasierte Anwendungen und Filetransfer

Mobilität im Allgemeinen hat von Natur aus einen negativen Effekt auf Webbrowsing. Bei der Nutzung eines Protokolls wie HTTP 1.0 wird eine kurzzeitige TCP-Verbindung zwischen dem Webserver und dem MN erstellt. Dabei ist die Wahrscheinlichkeit, daß ein Handoff genau in diese Verbindung fällt, relativ gering und auch von minimaler Auswirkung. Ein Nebeneffekt von Mobilität jedoch ist das Tunneln zwischen HA und dem fremden Netzwerk, wie es in Mobile IP definiert ist. Wenn eine Webseite übertragen wird, benutzt das erste TCP-Paket normaleweise die gesamte MTU. Durch IP-in-IP Kapselung wird die zulässige Größe des MTU-Feldes überschritten. Wenn das Paket nicht fragmentiert werden darf, führt dies zu einem ICMP Fehler und zusätzlicher Verzögerung bei jeder vergleichbaren TCP-Verbindung. Das passiert nicht, wenn sich der MN in seinem Heimatnetz bewegt, dessen Protokoll Hawaii oder CIP ist, weil sie im Gegensatz zu Mobile IP und HMIP ohne Tunneln auskommen.

Im Folgenden sollen die Auswirkungen regelmässiger Handoffs auf TCP-Verbindungen, z.B. durch Filetransfers geprüft werden. Dabei simulieren wir zwischen 0 und 30 Handoffs pro Minute, indem der MN zwischen AP_3 und AP_4 wechselt. Darauf basierend wird ein lang dauernder Download vom CN zum MN betrachtet. Für bessere Simulationsergebnisse gehen wir davon aus, daß das Internet (Abb. 4.9) kein Flaschenhals ist, sondern ebenfalls eine schnelle Netzwerkverbindung darstellt. Das Ergebnis des Versuchs (siehe 4.12) zeigt hier einen Vorteil von Paketverlust-minimierenden Ansätzen. CIP Semisoft und Hawaii MSF liefern vergleichsweise gute Ergebnisse, weil die möglicherweise falsche Reihenfolge und Verzögerung der Pakete auf höheren Schichten problemlos bearbeitet wird. CIP Hard Handoff und Hawaii UNF hingegen nehmen Paketverluste in Kauf, was in einem Abfall des Durchsatzes von fast 25% resultiert, wenn der AP alle 2 Sekunden gewechselt wird.

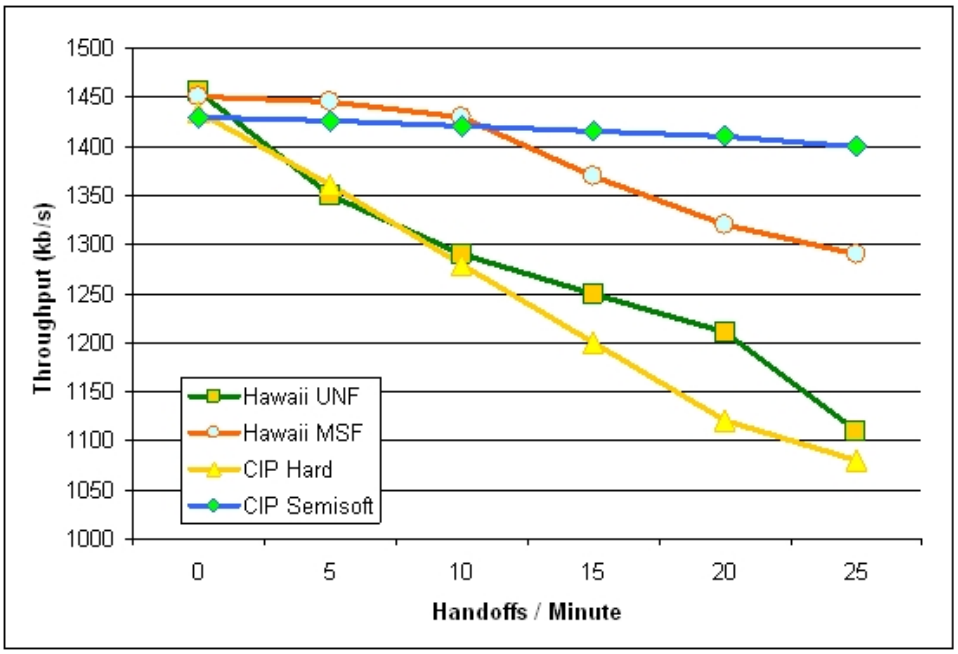

Abbildung 4.12: TCP Durchsatz auf Anwendungsschicht, aus [8]

4.6 Fazit

Diese Arbeit soll und kann keinen bewertenden Vergleich zwischen den verschiedenen Protokollen geben. Insbesondere im vorigen Kapitel wurde herausgestellt, daß sich die Protokolle und ihre einzelnen Schemata sehr unterschiedlich verhalten. Das größte Problem liegt im Einzelfall darin, den Bedarf an verlust- oder verzögerungsfreien Verbindungen zu bewerten und gegeneinander abzugrenzen. In der heutigen Situation werden IP-basierte Netzwerke für eine Vielzahl von Kommunikationssystemen genutzt, mit ganz unterschiedlichen Ansprüchen. Dabei seinen hier nur nochmal VoIP-Verbindungen erwähnt, die geringe Paketverluste zwar verschmerzen, aber Verzögerungen kaum verkraften können, weil sie dann für kommerzielle Anwendungen nicht konkurrenzfähig sind, wenn sie auch noch so günstig sein mögen. Der mobile Internetnutzer hat diese Ansprüche nicht. Gleichzeitig

hat es herausgestellt, daß die Protokolle Hawaii und Cellular IP in der Simulation gewisse Vorteile gegenüber dem Hierarchical Mobile IP Protokoll hatten. Sie haben jedoch den gravierenden Nachteil, daß sie nicht ohne das Ersetzen aller Router einer Domäne, also nur mit großem Aufwand, in bestehende Netze integriert werden können, während HMIP mit der Verwendung von Tunneln keine solchen Probleme kennt. Was sehr viel einfacher zu bewerten ist, sind die Ideen, die den Strategien und deren Umsetzung zugrunde liegen. Das sind insbesondere bei den „Konzeptstudien" Fast und Proactive Handoff Überlegungen, die Grenzen zwischen den OSI-Schichten 2 und 3 zu lockern. Es wurde zweifellos herausgestellt, daß diese Ansätze durchaus große Effekte in der Optimierung von Handoffverzögerungen haben. Das Aufbrechen des ISO/OSI-Schichtenmodells wird derzeit an vielen Anwendungsgebieten der Kommunikation diskutiert, weil es durch seine strikte Trennung der Schichten zahlreiche Optimierungsmöglichkeiten in der Kommunikation verhindert, gleichzeitig aber auch durch jahrelangen Bestand seine Existenzberechtigung hat. Ein Ansatz, der in diesem Zusammenhang geäußert wurde, ist die Definition einer „Open Radio API", um das heterogene Feld bestehender und zukünftiger Kommunikationsmedien zu vereinheitlichen. Dieser Ansatz wird vermutlich schwierig zu realisieren sein.

Ein weiterer, bereits mit der mobilen Telekommunikation eingeführter Begriff ist Paging. Durch die Einführung eines passiven Zustands kann die Frequenz der Location Updates reduziert werden. So kann Energie gespart werden, indem nur noch der L2-Stack regelmäßig auf wichtige Pakete untersucht wird und nur im Bedarfsfall gesendet werden muß. Gleichzeitig wird damit der Kommunikationsaufwand im Netz kaum sinken, da im Durchschnitt die eingesparte Kommunikation innerhalb des Netzes durch Broadcasting während des Suchens eines MN ausgeglichen wird. Je nach Anwendungsgebiet ist dieser Aufwand womöglich unnötig hoch. Bei einem Notebook beispielsweise verbraucht regelmäßige Kommunikation nur einen Bruchteil der zur Verfügung stehenden Energie gegenüber Komponenten wie Festplatte, Display und dem Prozessor selbst. Ist das Gerät jedoch auf Standby, sind in Zukunft durchaus Modelle denkbar, in denen der MN seine Position aktualisieren sollte, um Dienste wie WakeOnLAN für IP-basierte Kommunikation nutzen zu können. Dabei würde stetige Kommunikation einen sehr viel höheren Anteil am Energieverbrauch des MNs - ähnlich zu Mobiltelefonen - darstellen, und Paging eine zunehmend wichtige Rolle spielen.

Durch die Überlegungen zu Mobilität in IP-basierten Netzen ist man zwangsläufig zu der Einführung sogenannter „All-IP" Netze gekommen. Diese haben zur Folge, daß die bekannten Routingkonzepte durch neue, in jedem Fall kompliziertere Strategien ersetzt oder erweitert werden müssen, wie es in den einzelnen Protokollen beschrieben wurde. Die sich daraus ergebenden Routing Tabellen müssen riesige Datenmengen effizient verarbeiten und insbesondere aktualisieren können. Beispielweise müssen die Gateways der jeweiligen Netze stets Einträge für jeden einzelnen MN in der Hierarchie verwalten. Ob sich diese Ansätze auch in große Bereiche der heutigen Netztopologie integrieren lassen, wird abzuwarten sein.

Letztendlich kann man jedoch feststellen, daß unter dem breiten bestehender Konzeptstudien und Protokolle zahlreiche Ansätze gefunden wurden, um den Problemen im Zusammenhang mit Mobilität zu begegnen. Eine Gewichtung der Ergebnisse und absolute Bewertung der Protokollfähigkeiten wäre die Aufgabe einer weiteren Arbeit, die einer

Ausrichtung auf ein detailliert vorgegebenes Anwendungsgebiet bedarf.

Abkürzungsverzeichnis

CIP Das Cellular IP Protokoll

CIPR Cellular IP Router

CN Correspondent Node

COA Care Of Address

CR Crossover Node

FA Foreign Agent (Mobile IP, HMIP)

FN Foreign Network

GFA Gateway Foreign Agent (HMIP)

HA Home Agent (Mobile IP)

HMIP Hierarchical Mobile IP

IMN Idle Mobile Node (Paging)

MN Mobile Node

MRP Mobile Routing Point

PA Paging Area

PFA Paging Foreign Agent

SHRT Strong Handoff Radio Trigger

WIPPOA Wireless IP Point Of Attachment

Literaturverzeichnis

[1] Charles E. Perkins: Mobile Networking Through Mobile IP, Internet Tutorial http://www.computer.org/internet/v2n1/perkins.htm

[2] Andrew T. Campbell, Javier Gomez-Castellanos: IP Micro-Mobility Protocols, ACM SIGMOBILE, Oktober 2001. Aus Mobile Computer and Communication Review, Vol. 4, No. 4 pp. 45-53

[3] Andrew T. Campbell, Javier Gomez-Castellanos et al.: Performance of Cellular IP access networks, Center for Telecommunications Research, Columbia University, Januar 2000. Technical Report

[4] Pierre Reinbold, University of Namur: IP Micro-Mobility Protocols, 2003, IEEE Communications Surveys and Tutorials, www.comsoc.org/pubs/surveys

[5] Charles E. Perkins: IP Mobility Support for IPv4, Januar 2002

[6] A. Campbell, J. Gomez et al.: Internet Micromobility, COMET Group, Columbia University, New York, USA, Results from the Cellular IP Project, Journal of High Speed Networks No. 11, pp. 177-198, IOS Press 2002

[7] Ramachandran Ramjee, Member of IEEE, et al.: HAWAII: A Domain-Based Approach for Supporting Mobility in Wide-Area Wireless Networks, IEEE/ACM Transactions on Networking, Vol. 10, No 3, Juni 2002

[8] Andrew T. Campbell, Javier Gomez et al., Columbia University: Comparison of IP Micromobility Protocols, Februar 2002, IEEE Wireless Communications

[9] Hitesh Tewari et al.: Lightweight AAA for Cellular IP NTRG, Computer Science Department, Trinity College Dublin, Ireland

[10] E. Gustafsson, Annika Jonsson, Charles E. Perkins: Mobile IPv4 Regional Registration, IETF Mobile IP Working Group, November 2003, draft-ietf-mobileip-reg-tunnel-08.txt

[11] H. Haverinen, J. Malinen: Mobile IP Regional Paging, IETF Mobile IP Working Group, Juni 2000, draft-haverinen-mobileip-reg-paging-00.txt

[12] Karim El Malki, Hesham Soliman: Simultaneous Bindings for Mobile IPv6 Fast Handovers, Oktober 2003, draft-elmalki-mobileip-bicasting-v6-05.txt

[13] P.Calhoun et al.: Agent Assisted Hand-off, November 2000, draft-ietf-mobileip-proactive-fa-03.txt

[14] Edward C. Perkins: IP Mobility Support for IPv4, Internet RFC, RFC 3220 Januar 2002

[15] Praktikum Mobile Systeme, Institut für Informationstechnische Systeme (IIS), Fakultät für Informatik, Universität der Bundeswehr, FT 2003

[16] S. Y. Wang, H. Kung: A simple methodology for constructing an extensible and high-fidelity TCP/IP Simulator, IEEE INFOCOM 1999, pp. 1134-1143, 1999

[17] Columbia IP Micromobility Software, Columbia University, New York, http://comet.columbia.edu/micromobility

[18] . Yokota et al.: Link Layer Assisted Mobile IP Fast Handoff Method over Wireless LAN Networks, KDDI R&d Laboratories, Inc. Ohara Japan

[19] Jochen Schiller: Mobilkommunikation, Addison-Wesley Verlag, 2000, ISBN 3–8273–1578–6

Kapitel 5

Quality of Service in Wireless Local Area Networks

Andreas Fischer

Kabellose Netzwerke sind in verschiedenen Varianten verfügbar. Nun ist es für die Planung und den Vergleich von solchen Netzen notwendig, bestimmte Kriterien festzulegen. Diese werden im allgemeinen mit Quality of Service zusammengefasst. Aufgrund der besonderen Eigenschaften sind in der drahtlosen Umgebung die Mechanismen auf kabelgebundenen Netzen nur eingeschränkt anwendbar. Die grundsätzlichen Eigenschaften, mit denen die Netze beschrieben werden, sind weitgehend zu denen im Kabelnetz vergleichbar. Die Mechanismen, die in kabellosen Netzen zur Anwendung kommen, werden in diesem Abschnitt näher betrachtet. Dabei ist die Arbeit in zwei Bereiche unterteilt: im ersten Teil wird beschrieben, wie eine Bewertung von existenten Netzen, egal welcher Art, möglich ist. Der Zweite Absachnitt beschäftigt sich dann mit den Umsetzungen von Eigenschaften bezüglich Quality of Service in vorgegebenen Standards wie 802.11.

Inhaltsverzeichnis

5.1 Einleitung 107
5.1.1 Einordnung und Bedeutung von Quality of Service 107
5.1.2 Betrachtung anhand von verschiedenen Metriken 107
5.2 QoS nach IEEE 802.1 p 108
5.3 Betrachtung von QoS mittels RTFM 109
5.3.1 Traffic Meter 110
5.3.2 Meter Reader 112
5.3.3 Analysis Application 112
5.3.4 Manager 112
5.4 Unterstützung von QoS in verschiedenen Standards 113
5.4.1 Wireless Local Area Networks nach IEEE 802.11 113
5.4.2 Virtual Bridged Local Area Networks nach IEEE 802.1 Q 117
5.5 Zusammenfassung 118

5.1 Einleitung

In diesem Seminar wird die Umsetzung von Quality of Service (im Folgenden QoS)in einer Wireless Local Area Network Umgebung betrachtet. Dazu werden zwei grundsätzlich verschieden Ansätze verfolgt. Im ersten Teil der Arbeit wird die Unterstützung von QoS innerhalb der standartisierten Netze nach IEEE 802.11 betrachtet. Hierbei wird der Vergleich zu anderen Netzen gezogen. Der Zweite Teil der Arbeit beschäftigt sich dann mit der Betrachtung von QoS in einem vorgegebenen Netz. Dazu wird der Ansatz des Realtime Traffic Flow Measurement näher verfolgt. Dieser beschreibt eine Betrachtung des Datenverkehrs unabhängig von der Schicht aus dem ISO/OSI-Referenzmodell.

5.1.1 Einordnung und Bedeutung von Quality of Service

Der Begriff "Quality of Service", deutsch meist "Dienstgüte", wird in der Literatur teilweise widersprüchlich interpretiert. Es seien hier drei gängige Definitionen beschrieben: QoS definiert als die Zusammenfassung verschiedener Kriterien, die die Güte eines Netzwerk-Dienstes bezüglich Zeit und Zuverlässigkeit betrachten, Garantiewerte der Leistungsfähigkeit eines Datennetzes bezüglich Zeit und Zuverlässigkeit oder auch: Anwendungen eine bestimmte Güte des Netzes bezogen auf Zeit und Zuverlässigkeit garantieren zu können.

Im weiteren Verlauf dieser Arbeit wählen wir eine etwas weiter gefasste Definition des Begriffes: QoS beschreibt Anforderungen an ein Kommunikationssystem aus einer bestimmten Sicht. Damit schließen wir auch andere Aspekte als Zuverlässigkeit und Zeit ein. Dies könnte etwa Bandbreite oder Sicherheit sein. Mit der Sicht sei beschrieben, das die Anforderungen an ein Kommunikationsnetzwerk je nach Betrachtungsweise sehr unterschiedlich sein können. Darauf werde ich mit Beispielen im folgenden Gliederungspunkt weiter eingehen.

5.1.2 Betrachtung anhand von verschiedenen Metriken

Definieren wir jetzt also gemäß unserer Definition verschiedene Metriken der Anforderungen. Angemerkt sei, dass hier nur ein exemplarischer Ausschnitt dargestellt werden kann.

Die Sicht wird hier in drei verschiedene Klassen geteilt: den Provider, den Dienstanbieter und den Verbraucher. Dabei wir der Netzbetreiber als Provider gesehen. Für ihn sind Ziele wie hohe Auslastung des Netzes und geringe Betriebskosten interessant. Als Dienstanbieter ist ein möglicher Verkäufer gedacht, dem der Provider eine bestimmte Funktionalität bereitstellt, über die der Kundenkontakt und die Übertragung der Leistung erfolgt. Für diesen Teilnehmer an der Kommunikation sind mögliche Garantien bezüglich Sicherheit, Vertraulichkeit, Integrität der Daten, Zeit besonders wichtig, um seine Dienstleistung anzubieten.

Beschreiben wir zuerst die Kapazität von Netzwerken. Dies ist, eine Metrik, die besonders für die Entwicklung und für den Betrieb von Netzwerken interessant ist. Dazu kann man

etwa die Kapazität von Netzwerken bezüglich der möglichen Anzahl der Knoten oder der Endsysteme betrachten. Dazu besteht oft eine Abhängigkeit von den Datenmengen, die über das Netz transportiert werden. Somit kann unter dem Begriff Kapazität auch die Bandbreite zu verstehen sein, die das entsprechende Netz besitzt. Weiterhin kann für den Netzbetreiber die mögliche Ausdehnung des Netzes einen Einfluss auf die Anwendungsmöglichkeiten haben.

Eine Metrik, insbesondere in Bezug auf Sicherheit, ist die Fehlerrate. Auch diese ist für den Provider von Bedeutung. Die Fehlerrate kann Einfluss auf die Gesamtdatenmenge haben. So wird unter Umständen eine Neuübertragung von fehlerhaften Paketen angeregt werden. Dies ist auch für einen Dienstanbieter ein entscheidendes Merkmal, um zwischen verschiedenen Providern zu wählen. Weiterhin können Applikationen den Fehlerraten angepasst werden, um auf höherer Ebene eine fehlerlose Übertragung zu ermöglichen. So ist die Fehlerrate also für Dienstanbieter und Provider ein wichtiges Merkmal zur Bewertung von Kommunikationsnetzwerken.

Eine weitere Metrik, die besonders für übergeordnete Schichten nach dem ISO/OSI-Referenzmodell von Bedeutung ist, stellt sich in Form der Varianz der Verzögerungszeiten (englisch: Jitter) dar. Dieser Wert ist insbesondere für Real-Time-Anwendungen interessant. Entsprechende Garantien sind, wie bereits in Teil 3 dieses Seminars beschrieben, besonders in drahtlosen Umgebungen nur schwer zu realisieren.

Dazu kommen einige Metriken, die für den Provider weniger interessant sind. Dazu gehören etwa Verzögerungs- und Antwortzeiten. Das heißt aber nicht, dass entsprechende Werte für den Netzbetrieb unwichtig sind. Kunde und Dienstanbieter sind an einer definierten Leistung des Netzes in Bezug auf diese Metriken interessiert. So macht es für den Betreiber einer Internetseite einen erheblichen Unterschied, ob http-Anfragen in kurzer Zeit beantwortet werden können oder ob sich für den Kunden eine große Verzögerung ergibt.

Ein erheblicher Teil der Antwortzeit für den Kunden ergibt sich aus der Latenzzeit der Übertragung im Netz. Daher ist es möglich, das Servererweiterungen nur einen geringen Einfluss auf die Antwortzeit hat. Weitere Möglichkeiten für die Bewertung von Netzwerken sind Aufbauzeit und Haltezeit für eine Verbindung. Auch diese Metriken sind für Kunden und Dienstanbieter entscheidend, um auf höheren Ebenen damit umzugehen.

5.2 QoS nach IEEE 802.1 p

Für eine Netzstruktur nach den Spezifikationen der IEEE hat das Gremium einen Standard für QoS in IEEE 802.1 p gesetzt. Hier werden QoS die folgenden Metriken zugeordnet: Service Availebility beschreibt die Verfügbarkeit der Verbindungsstruktur für ein Endgerät. Dabei wird die Rate zwischen Verfügbarkeit und keiner Verfügbarkeit gemessen. Um die Verfügbarkeit zu erhöhen, soll eine automatische Wiederherstellung der Konfiguration implementiert werden.

Weiterhin wird der Paketverlust als Metrik für QoS beschrieben. Der Verlust von Paketen auf der Schicht 2 ist selten, aber da eine Übertragung nicht garantiert ist, wird dieser

Punkt in die QoS-Betrachtungen auf der Ebene 2 des ISO/OSI-Referenzmodells einbezogen. Der Verlust von Datenpaketen kann unter anderem an der begrenzten Kapazität der Pufferspeicher in den Netzkomponenten liegen. Weiterhin kann er durch eine lange Laufzeit bedingt sein. Wenn die Laufzeit das TTL-Feld des Frames übersteigt, wird dieser verworfen.
In der Spezifikation wird auch die Reihenfolge von Datenpaketen betrachtet. Wird diese also im Netz verändert, kann es zu einer Veränderung der Information kommen. Daher ist die Reihenfolge auch für QoS interessant.
Innerhalb der Schicht 2 wird kein Datenpaket dupliziert, da die Vorgaben dieses Standards sich nicht nur auf die MAC-Ebene beziehen, ist die Vervielfältigung von Frames hier mit aufgeführt, wird im betrachteten Zusammenhang allerdings nicht weiter ausgeführt.
Wie in fast allen Definitionen von QoS wird auch in dieser die Verzögerungszeit betrachtet. Damit ist in diesem Zusammenhang die Zeit zwischen der Anfrage und der erfolgreichen Übertragung der entsprechenden Antwort zu verstehen.
Wie bereits beschrieben haben die Datenpakete eine Lebenszeit. Wird diese überschritten, wird der Frame nicht weiter übertragen. Das hat einen Einfluss auf verschieden andere Metriken von QoS.
Die Frames auf Schicht 2 haben eine Obergrenze bezüglich ihrer Größe. Damit ergibt sich ebenfalls eine maximale Größe der Daten, die innerhalb eines Paketes übertragen werden können.
Neben diesen Metriken sind in den Spezifikationen eine Benutzerpriorisierung und der Durchsatz definiert. Diese sind auf den Ebenen 1 und 2 jedoch nicht umgesetzt.

5.3 Betrachtung von QoS mittels RTFM

Die Analyse von Netzen unter den bisher beschriebenen Metriken ist sehr komplex. Im Folgenden wird eine Möglichkeit zur Bewertung von Netzverhalten beschrieben. Hier ist kein Schwerpunkt auf die Metriken von QoS gesetzt, aber durch den freien Ansatz können insbesondere solche betrachtet werden.

Schauen wir also auf die grundlegenden Eigenschaften von Realtime Traffic Flow Measurement. Dieser Standard wurde innerhalb einer Arbeitsgruppe der Internet Engineering Task Force entwickelt. Ziel ist ein applikationsorientiertes Interface zur Bewertung von Datenströmen. Dazu muss ein Datenstrom (im Folgenden Flow) definiert werden. Im Zusammenhang der IETF wird eine relativ freie Definition genutzt. Ein Flow besteht aus verschiedenen Paketen, die logisch zusammenhängen. Wie der Zusammenhang dabei zu verstehen ist, wird nicht spezifiziert. Denken wir an die QoS-Metriken, sind hier etwa zeitliche Zusammenhänge denkbar. Weiterhin können im allgemeinen Umfeld Quell- und Zieladressen betrachtet werden. Es sind aber auch ganz andere Zusammenhänge, wie Größe oder Priorität denkbar.

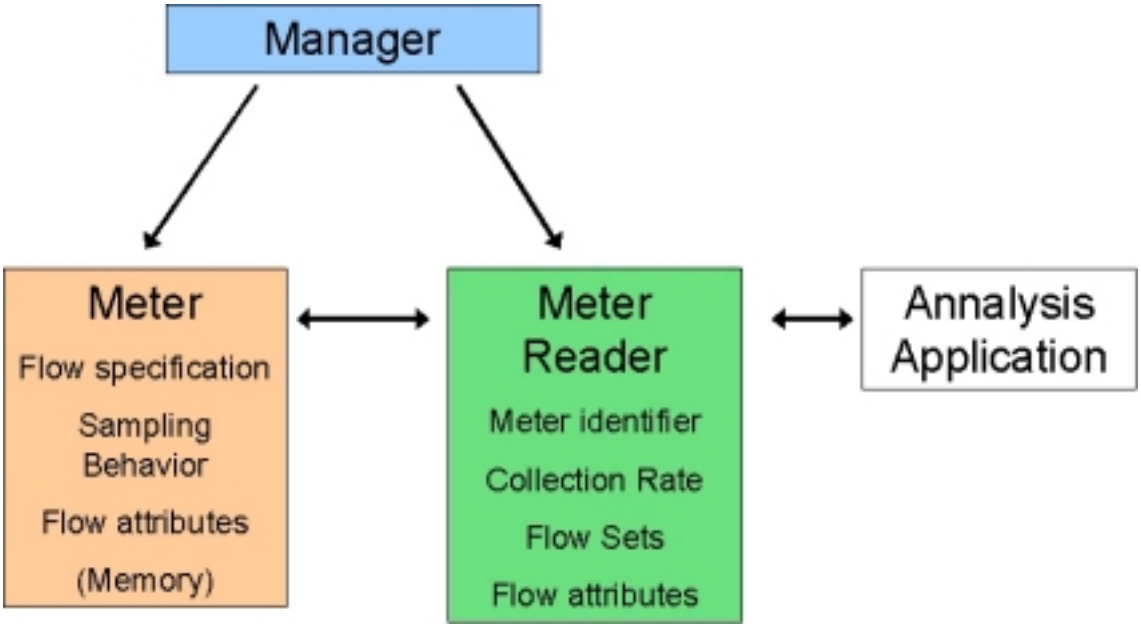

Abbildung 5.1: Aufbau von RTFM

Die Architektur von RTFM gliedert sich in vier Teile, die interagieren. Die Funktionseinheit, die den eigentlichen Datenverkehr überwacht, ist der Traffic Meter. Vom Meter Reader werden die gesammmelt und geordnet. Die strukturierten Daten des Meter Reader werden zur Analysis Application übertragen und hier ausgewertet. Zur Administration des Ablaufes ist der Manager unabhängig von allen anderen Einheiten installiert.

5.3.1 Traffic Meter

Der Traffic Meter ist eine Einheit, die Daten zum Betrieb an einem bestimmten Punkt innerhalb des Netzwerkes, dem Metering Point, sammelt. Er kann in verschiedenen Formen ausgeführt sein: wird der Datenstrom in einem Host des Netzwerkes betrachtet, kann die Funktionalität in einem Programmteil der Software abgebildet sein. Hier wird dann der Header der Pakete betrachtet und nach gewissen Eigenschaften ausgewertet. Wird ein System mit verschiedenen Netzzugängen bewertet, so wird eine Struktur verwendet, die die Pakete an allen Zugängen analysieren kann. Damit kann jeder Zusammenhang zwischen verschiedenen Zugangspunkten betrachtet werden. Ein Spezialfall dieses Szenarios wird in Knoten verwendet, die Pakete weiterleiten; so etwa Switch oder Router.

Der Traffic Meter benötigt zur Auswertung des Datenverkehrs im Netzwerk ein bestimmtes Wissen. So müssen ihm sowohl die Flow-Spezifikation als auch die Flow-Attribute bekannt sein. Dazu kommt ein bestimmtes Verhalten bei der Aufzeichnung der Daten. Dabei könnten etwa nur ausgewählte Pakete betrachtet werden. So ist es möglich, jedes n-te Paket oder Pakete ab einer bestimmten Größe in die Bewertung einzubeziehen. Diese Anforderungen an den Traffic Meter erfordern einen Speicher sowie Rechenkapazität.

Die Arbeitsweise eines Traffic Meters stellt sich wie folgt dar: Ein Packet Processor empfängt die Pakete. Dabei betrachtet er den Paketkopf und liest den Match Key aus. Durch

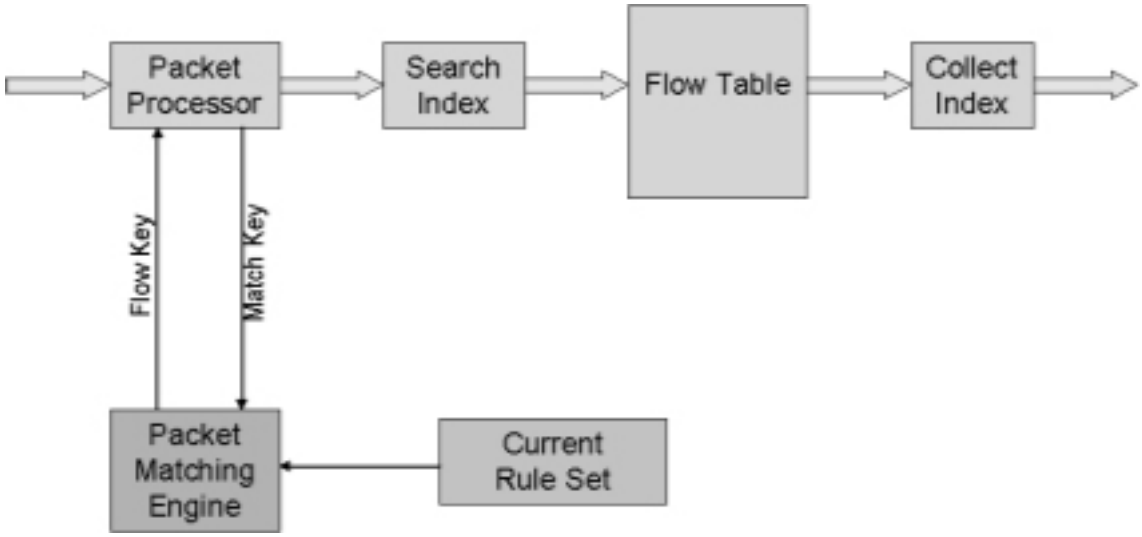

Abbildung 5.2: Paketverarbeitung nach RTFM

den Match Key wird ein Paket eindeutig bestimmt. Der Match Key wir dann an die Packet Matching Eingine übergeben. Hier liegen die einzelnen Flow-Spezifikationen vor. Anhand des Match Key und der Spezifikationen wird der Flow Key bestimmt. Dieser ordnet das Paket einem Flow zu. Die gesammelten Informationen werden jetzt an den Meter Reader übertragen. Unter dem Flow Key werden die entsprechenden Paketinformationen vom Packet Processor in den Flow-Table geschrieben.

Ein Eintrag in den Flow Table enthält midestens folgende Daten:

- Adressen von Absender und Addressat, welche schichtunabhängig gestaltet sind
- Zeitstempel des ersten und des letzten Paketvorkommens des Flow.
- Zähler für Hin- und Rückrichtung des zum Flow gehörigen Verkehrs
- weitere frei definierbare Attribute

Die einzelnen Datenströme können die können nach folgenden Zuständen unterschieden werden:

- Inaktiv, wenn der Flow Record nicht vom Traffic Meter genutz wird
- Current, wenn der entsprechende Flow vom Meter genutzt wird, um Pakete zu analysieren
- Idle, wenn der Record zwar vom Meter genutzt wird, aber für eine spezifizierte Zeit keine Pakete zum entsprechenden Flow zugeordnet wurden

  end itemize Dazu können hier weitere Funktionalitäten existent sein.

5.3.2 Meter Reader

Die so gesammelten Daten werden dann über ein festes Interface zu vom Manager definierten Zeiten an den Meter Reader übertragen. Hier erfolgt die Zusammenstellung in zu einem Flow Data File. Dazu werden im Standard keine weiteren Ausführungen gemacht.

5.3.3 Analysis Application

Die vom Meter Reader gesammelten Daten werden an die Analysis Application übergeben. Dazu macht der Standard keine weiteren Ausführungen. Es ist nach den Vorgaben auch möglich, mit der Applikation in das Netzwerk einzugreifen. So wäre ein Eingriff in die Netzwerktopologie denkbar. Damit würde eine dynamische Netzwerkstruktur entstehen. Dies könnte insbesondere für Multiprozessorsysteme interessant sein.

5.3.4 Manager

Dazu kommt der Manager. Seine Funktionalität besteht in der Konfiguration von Traffic Meter und Meter Reader. Dies geschieht mittels folgender Funktionen:

- Kontrollfunktionen zwischen Manager und Meter
  - Download Rule Set, um Rule Sets beliebiger Art auf den Meter zu übertragen
  - Specify Meter Task, um genau einen Rule Set auszuwählen und zur Anwendung zu bringen
  - Set high water mark, um den Flow Table in der Größe zu begrenzen
  - Set flow termination parameters, um das Ende des Flows zu spezifizieren
- Kontrollfunktionen zwischen Manager und Meter Reader
  - Identification between Manager and Meter reader
  - Reporting Interval Control, um den Abstand zwischen zwei Berichten vom Meter zum Meter Reader zu bestimmen
  - Granularity Control, um die Granularität der aufzuzeichnenden Daten zu bestimmen
  - Flow Lifetime Control, um zu bestimmen, wie lange die Daten zu einem inaktiven Flow nachgehalten werden

Metriken, die dem Flow nach RTFM zugeordnet werden müssen, sind Lebenszeit, Zähler und Zeitstempel. Dabei sind die Zähler als Rolling Counter auszulegen. Einfache Zähler wären mit der Instanziierung des Flow zu beginnen. Dann wird der Zähler einfach erhöht, solange der Flow existent ist. Der Rolling Counter wird, sobald die Informationen vom Meter Reader an die Analysis Application weitergegeben wird, zurückgesetzt. Durch dieses System wird das Verfahren weniger fehleranfällig. Sollten bestimmte Einheiten der gesammelten Informationen verloren gehen, wird die

Analyse von dem Teil einfach ausgeblendet, die Analyse des gesamten Flow bleibt davon nahezu unbeeinträchtigt. Neben den vorgegebenen Parametern kann man frei weitere definieren. So ist es möglich, etwa die Stationen von Paketen aufzuzeichnen. Damit ist leicht ersichtlich, dass es mit dem RTFM-Ansatz die Möglichkeit bietet, den Datenverkehr in bestimmten Netzen anhand von QoS-Metriken zu analysieren. Damit ist ein Hilfsmittel zu Bewertung von QoS gegeben.

5.4 Unterstützung von QoS in verschiedenen Standards

5.4.1 Wireless Local Area Networks nach IEEE 802.11

Betrachten wir jetzt die Umgebung nach der Spezifikation 802.11 der IEEE. Dieser Standard spezifiziert eine Wireless Local Area Network Umgebung. Entsprechende Geräte sind in vielfältigen Varianten am Markt. Fast alle Geräte aus dem Bereich der Wireless Networks entsprechen mindestens einem Standard der 802.11 Gruppe. Die IEEE hat in verschiedenen Varianten der Entwicklung der Funkübertragung Rechnung getragen. So sind verschiedene Netzwerkspezifikationen entstanden, die im Wesentlichen unterschiedliche Übertragungsraten ermöglichen. Der derzeit wohl gebräuchlichste Standard ist IEEE 802.11 b mit einer maximalen Übertragungsrate von 11 Mbit/s. Andere Standards, wie etwa 802.11 a und 802.11 g haben eine maximale Übertragungsrate von 54 Mbit/s. Sie sind teilweise kompatibel zu 802.11 b mit entsprechend geringeren Datenraten. Die Reichweite der Geräte im Standard 802.11 b beträgt je nach Umgebung etwa 25 bis 50 Meter.

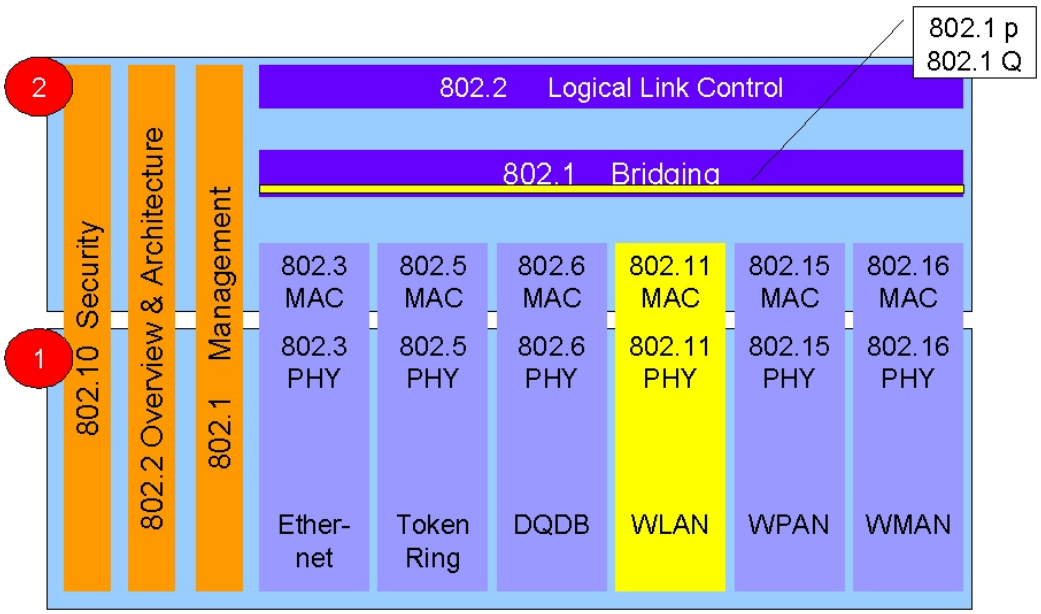

Abbildung 5.3: Einordnung von WLAN nach 802.11

Die Spezifikationen aller Standards 802.11 erstrecken sich über die Schichten eins und zwei des ISO/OSI-Referenzmodells. In der Gesamtarchitektur, die durch die

IEEE beschrieben ist, bietet 802.11 eine Schnittstelle nach oben an. Parallel zu der WLAN-Umgebung können auch Netzwerke wie etwa Ethernet oder Token Ring eingesetz werden. Dur die weitgehende Transparenz der Schnittstelle ist hier die Möglichkeit gegeben, diese Netze in einer gemeinsamen Umgebung anzuwenden, so wie es im Internet realisiert ist. Auf dieser Schnittstelle arbeiten dann verschiedene Protokolle des Bereiches Bridging und Logical Link Control.

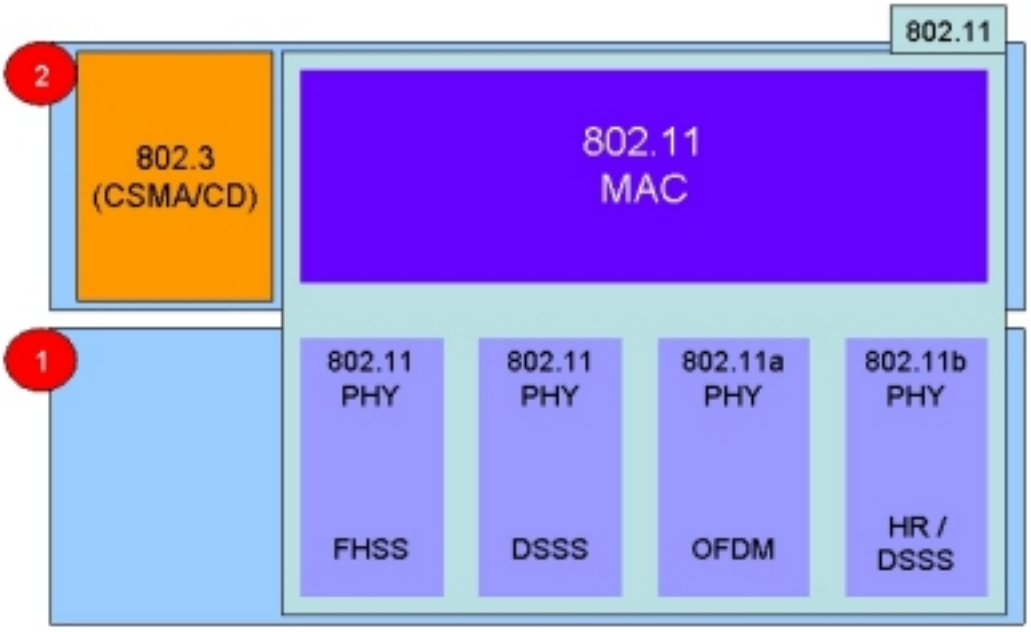

Abbildung 5.4: Gliederung der Bitübertragungsschicht nach 802.11

Innerhalb der Standardisierung 802.11 können auf der Schicht eins vier verschieden Verfahren genutzt werden. Alle bieten unter unterschiedlichen Paketrahmen die gleiche Funktionalität an, so dass dann auf der Schicht zwei nur ein einziges Paketformat existiert.

Im Folgenden wird beschrieben, welche einzelnen Felder innerhalb der Pakete der Übertragung zur Informationsgewinnung bezüglich QoS genutzt werden können. Betrachten wir zuerst den MAC-Frame auf Schicht zwei.

Der Paketkopf der MAC-Datenpakete hat eine feste Länge von 240 Bit. Er besteht aus einem 16 Bit langem Feld zur Paketkontrolle. Dazu kommen jeweils 16 Bit Duration/ID und Sequenzkontrolle. Mit dem Feld Duration wird die Dauer der Paketübertragung übertragen. Dies ist notwendig, da die Größe des zu übertragenden Datums nicht festgelegt ist. Mit den 16 Bit Sequenzkontrolle ist es möglich, 65536 Datenpakete in einer Sequenz zu übertragen. Damit ergeben sich maximal etwas mehr als 1,5 GB in einer Sequenz.

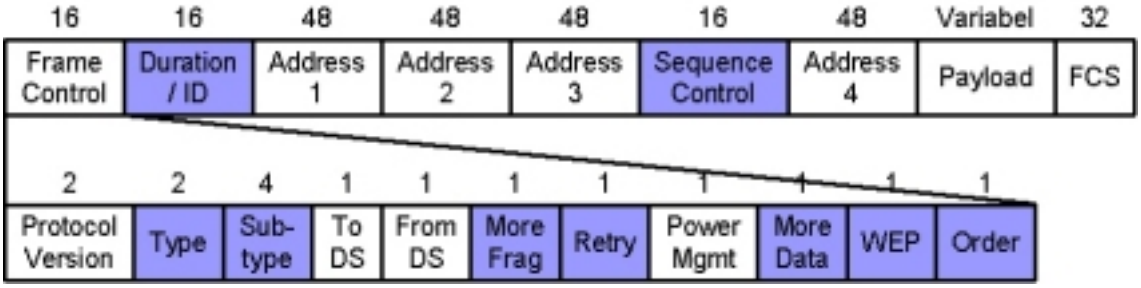

Abbildung 5.5: MAC–Frame nach IEEE 802.11

Type und Subtype-Feld des Paketkontrollkopfes spezifizieren die Art des MAC-Frames. Es gibt Management-Frames, die zum Beispiel bei der Assoziierung von Stationen verwendet werden, Control-Frames wie RTS/CTS und Data-Frames.

Die anderen Flags im zwei Byte breiten Frame Control-Feld geben an, ob es sich um einen Frame zum oder vom Distribution System (DS) handelt, ob noch weitere Fragmente folgen, da die MSDU geteilt wurde, ob es sich um eine wiederholte Verschickung des Frames handelt, ob die sendende Station sich im Power-Saving-Modus befindet, ob sie weitere Daten versenden will und ob die Daten mittels WEP verschlüsselt sind und Fragmente in der richtigen Reihenfolge versandt werden.

Bei Control Frames sind alle 1 Bit-Flags mit Ausnahme des Power Management-Flags auf Null gesetzt. Der IEEE 802.11-MAC unterscheidet fünf verschiedene MAC-Adressen, die Basic Service Set Identification (BSSID), Destination Address (DA), Source Address (SA), Receiver Address (RA) und Transmitter Address (TA). Die BSSID ist in der Regel die MAC-Adresse des Access Points. Der AP ist auch die Station, an die Frames mit gesetztem To DS-Flag gesendet werden beziehungsweise die Station, die Frames mit gesetztem From DS-Flag versendet. Die Anzahl der verwendeten Adreßfelder sowie deren Inhalt (BSSID, DA, SA, RA oder TA), werden von den Flags To DS und From DS festgelegt. Sind beide Flags 1, werden nur die Adressen 1 bis 4 belegt. In allen anderen Fällen entfällt Adresse 4, so dass der Paketkopf nur 192 Bit lang ist. Sind beide Flags 0, so ist Adresse 1 die DA, Adresse 2 SA und Adresse 3 BSSID. Ist To DS 0 und From DS 1, so beschreibt Adresse 1 die DA, Adresse 2 BSSID und Adresse 3 die SA. Bei der Belegung To DS 1 und From DS 0 steht die BSSID in Adresse 1, die SA in Adresse 2 und die DA in Adresse 3. Sind beide Flags 1, werden wie beschrieben alle vier Adressen besetzt. Dazu ist Adresse 1 mit RA belegt, in Adresse 2 findet sich die TA. Adresse 3 und 4 beschreiben DA und SA.

Es werden neben Unicast-Adressen auch Multicast- und Broadcast-Adressen verwendet. Die Broadcast-Adresse lautet 111....111. Während Frames an eine Unicast-Adresse in der Regel vom Empfänger durch Versenden eines ACK bestätigt werden, geschieht dies bei Frames an eine Multicast- oder Broadcast-Adresse nicht, da es durch das zeitgleiche Versenden der ACK-Frames von allen Empfangsstationen zu einer Kollision kommen würde.

Das FCS-Feld enthält einen 32 Bit-CRC-Code über den gesamten Frame inklusive Frame Body. Ist ein empfangener Frame fehlerhaft, so wird es an diesem Feld erkannt. Fehlerhafte Frames werden nicht bestätigt.

Wesentlich für QoS sind die folgenden Informationen des Paketkopfes: Type und Subtype, More Fragmentation, Retry, More Data, WEP und Order aus dem Bereich Frame Control sowie Duration/ID und Sequence Control. Mit den Feldern Type und Subtype werden, wie bereits beschrieben, verschiedene Framearten bestimmt. Diese können im Netz unterschiedlich behandelt werden. So wird eine Bestätigung immer nur mit einem Short Interframe Space versandt. Damit ergibt sich eine höhere Priorität. Das Feld Retry beschreibt eine erneute Übertragung. an diesem Feld kann auch der Empfänger erkennen, dass ein vorheriger Übertragungsversuch fehlerhaft war.

Mit dem Feld More Data kann der Adressat des Paketes verifizieren, ob dieses Paket mit anderen in Zusammenhang steht oder der Inhalt isoliert betrachtet werden kann. Das Feld WEP signalisiert, dass das Nutzdatum mit Wired Equivalent Privacy, dem Sicherheitsprotokoll von 802.11, verschlüsselt ist.

Das Order-Feld wird gesetzt, wenn die Pakete in einer bestimmten Reihenfolge versandt wurden und die Nutzdaten so zu interpretieren sind. Duration/ID beschreibt die Länge der Frameübertragung. Damit ist es möglich, in Netzkomponenten bestimmte Schedulingalgorithmen für die Bearbeitung der Pakete zu implementieren. Damit kann man eine Priorisierung von Paketen anhand deren Länge vornehmen. Mit der Sequenzkontrolle ist es möglich, bestimmte Frames in einer Sequenz zu übertragen, so dass sie am Empfänger in die vorgegebene Reihenfolge zu bringen sind.

Kommen wir jetzt zu den Datenpaketen auf Schicht eins. Hier sind vier verschiedene Verfahren zur Datenübertragung spezifiziert: Frequency Hopping Spread Spectrum (FHSS), Direct Sequence Spread Spectrum (DSSS), Orthogonal Frequency Devision Multiplexing (OFDM) und High Rate Direct Sequence Sprad Spectrum (HR/DSSS).

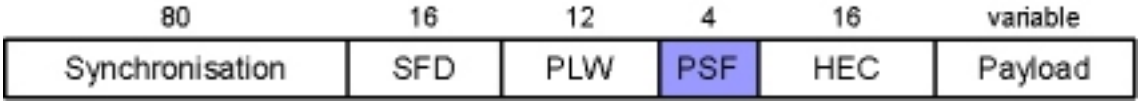

Abbildung 5.6: FHSS–Frame nach IEEE 802.11

Der FHSS-Frame besteht aus den Feldern Synchronisation, SFD, PLW, PSF und HEC. Der Paketkopf hat eine Gesamtlänge von 128 Bit. Die Felder Synchronisation und Start Frame Delimiter bestimmen die PLCP-Präambel. Der PLCP-Paketkopf enthält die Felder Payload Length Word, Payload Signalling Field und Header Error Check. Für QoS-Eigenschaften kommt hier nur die Nutzung des Feldes Payload Signalling Field in Frage, in dem die Datenrate der Übertragung des Nutzdatums bestimmt ist. Hierbei steht 0000 für 1 Mbit/s, die maximale Übertragungsrate beträgt 8,5 Mbit/s und wird mit 1111 gekennzeichnet.

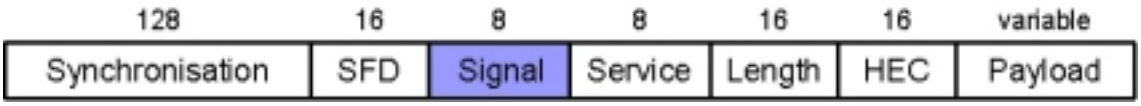

Abbildung 5.7: DSSS–Frame nach IEEE 802.11

Der DSSS-Header ist grundsätzlich 192 Bit lang, die sich auf die Felder Synchronisation, Start Frame Delimiter, Signal, Service, Length und Header Error Check verteilen. Die Datenrate der Nutzdaten wird hier im Feld Signal übertragen. Dabei sind vier verschieden Datenraten vorgegeben, die sich in Hexadezimalzahlen wie folgt wieder finden: 0A für 1 Mbit/s, 14 für 2 Mbit/s, 37 für 5,5 Mbit/s und 6E für 11 Mbit/s. Die anderen Felder haben für QoS-Eigenschaften keine Bedeutung.

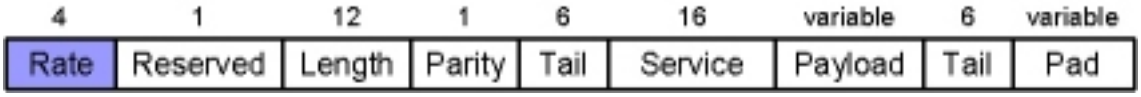

Abbildung 5.8: OFDM-Frame nach IEEE 802.11

Bei der OFDM-Übertragung ist der Paketkopf nur 40 Bit lang. Hier bestimmt das Feld Rate die Datenrate des Nutzdatums. Dazu steht die 3 für 54 Mbit/s, die 9 für 24 Mbit/s und die Zahl F für 9 Mbit/s. In den anderen Feldern werden Informationen übertragen, die für QoS ohne Bedeutung sind.

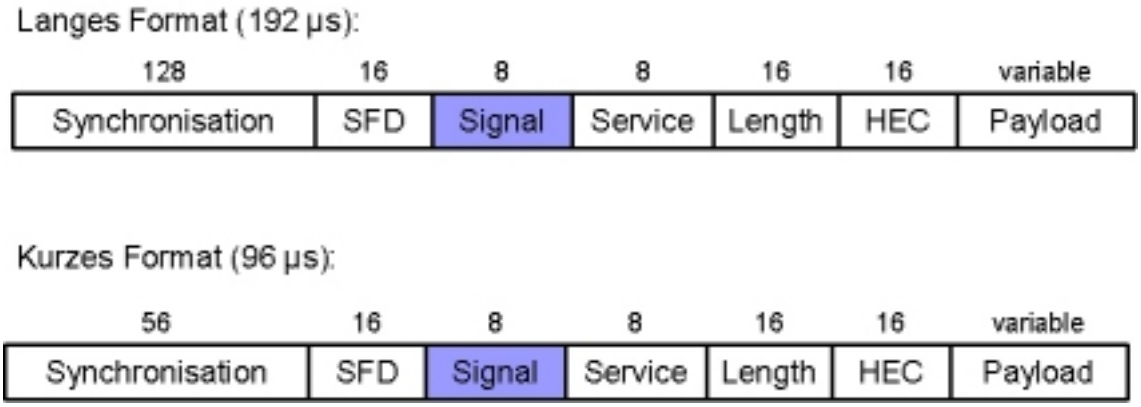

Abbildung 5.9: HR/DSSS–Frame nach IEEE 802.11

Den HR/DSSS-Frame gibt es mit zwei verschiedenen Paketköpfen, die sich nur in der Länge der Synchronisation unterscheiden. Mit einer kurzen Synchronisation ergibt sich eine Gesamtlänge des Header von 120 Bit, mit der langen Synchronisation 192 Bit. Auch in diesem Paket wird im Feld Signal die Datenrate übertragen. Die anderen Felder enthalten keine Informationen, die in Bezug auf QoS von Bedeutung sind.

5.4.2 Virtual Bridged Local Area Networks nach IEEE 802.1 Q

Das VLAN nach dem Standard der IEEE bezieht sich auf die Ebene zwei des ISO/OSI-Referenzmodells. Grundsätzlich werden hier drei verschiedene Paktarten unterschieden. Untagged Frames werden nicht direkt einer Gruppe von Teilnehmern zugewiesen. Die Zuordnung kann jedoch nach Kriterien wie etwa der Absenderadresse oder der Zieladresse erfolgen. Die zweite Klasse an Paketen sind die Priority Tagged Frames. Diese besitzen zwar eine Priorität, werden damit aber keinem VLAN zugeordnet. Dies geschieht, wie

auch bei den Untagged Frames, nach anderen Parametern, wie etwa Adressen. Die dritte Gruppe der Pakete sind die VLAN-tagged Frames. Diese werden nach dem entsprechenden Tag zu verschiedenen VLAN-Umgebungen zugeordnet. Hier werden im Netz feste VLAN implementiert, die nur über die Administration geändert werden. Dazu gehört dann auch die entsprechende Einstellung der Netzwerkelemente, wie zum Beispiel der Switches.

Damit die Pakete im Netz entsprechen behandelt werden, ergibt sich die Notwendigkeit für bestimmte Funktionalitäten der Elemente. Da der Standard auf Schicht zwei aufsetzt, sind hier die Vorgaben für Bridges zu definieren. Es werden verschieden Funktionen beschrieben, die in einer Bridge realisiert sein müssen, dass der Einsatz in einer VLAN-Umgebung sinnvoll ist.

Zu implementieren sind die folgenden Funktionen: Paketannahme, das Verwerfen von fehlerhaften Paketen nach dem Standard ISO/IEC 15802-3, das Verwerfen von Paketen, die keine Nutzdaten enthalten, Wiederherstellung der Benutzerpriorität im Bedarfsfall, Verwerfen von Paketen nach Vorgaben der Filterinformationen, Verwerfen von Paketen mit zu großem Nutzdatum, Paketweiterleitung an einen bestimmten Port, Auswahl nach Verkehrsklassen nach den Filterinformationen, Pufferung der Pakete nach Verkehrsklassen, Verwerfen von Paketen, um die maximale Verzögerungszeit der Bridge zu gewährleisten, Auswahl aus den gepufferten Datenpaketen zur Weiterleitung, Auswahl einer minimalen Benutzerpriorität, die bedient wird, Bearbeitung der Nutzdaten und Neuberechnung der Frame Check Sequence und Paketübertragung.

Die folgenden Funktionen sind für QoS-Eigenschaften der Netze von Bedeutung: Benutzerpriorisierung, Pufferung und Paketauswahl nach Verkehrsklassen, um so eine Priorität für bestimmten Verkehr zu erreichen, maximale Verzögerungszeit eines Paketes innerhalb der Bridge und Bestimmung der minimalen Priorität für weiterzuleitende Pakete.

Zusammenfassend kann man sagen, dass ein Virtual Bridged Local Area Network nach diesem Standard nur bestimmte Prioritäten für die Pakete festlegt. Diese werden dann in den einzelnen Bridges so ausgewertet, dass virtuelle Bereiche im Netz entstehen. Über die maximale Verzögerung einer Bridge wird eine feste Definition von Timeout-Siganlen möglich. Dabei ist es jedoch möglich, dass stark belastete Bridges eine hohe Verlustrate von Paketen aufweist.

5.5 Zusammenfassung

In Wireless Local Area Network – Umgebungen nach der IEEE 802.11 sind bezüglich QoS nur schwache Aussagen möglich. Der Standard bezieht sich nur auf die Ebenen eins und des ISO/OSI-Referenzmodells. Im Standard werden nur Informationen bezüglich der Reihenfolge der Pakete, der Vervielfältigung von Paketen und der maximalen Größe des Nutzdatums gemacht. Andere Informationen werden im Standard selbst nicht bereitgestellt.

Mit einer Umgebung nach dem Standard 802.1 Q “Virtual bridged LAN“auf einer drahtlosen Übertragung nach IEEE 802.11werden zusätzlich Benutzerpriorisierung und Aussagen über die maximale Verzögerungszeit in Netzwerkkomponenten möglich. Damit kön-

nen mehr Aussagen bezüglich QoS getroffen werden. Eine umfangreiche Auswertung aller Metriken ist mittels RTFM umzusetzen. Hier ist der weitgehend offene Ansatz hilfreich. Umgebungen auf Basis der Funkübertragung haben im Bereich QoS erhebliche Nachteile gegenüber kabelgebundenen Netzwerken. Die Entwicklung ist im Verhältnis noch in einem frühen Stadium. Dazu kommt, dass die technologischen Möglichkeiten in den bestehenden Standards nicht ausgenutzt werden. Eine Entwicklung von WLAN ist in naher Zukunft besonders im Bereich QoS zu erwarten.

Literaturverzeichnis

[1] Quality of Service on the MAC level,
http://www.tml.hut.fi/Opinnot/Tik-110.551/1999/papers/08IEEE802.1QosInMAC/qos.html

[2] IEEE 802.11 Bitübertragungsschicht
http://ivs.cs.uni-magdeburg.de/EuK/Lehre/Wintersemester_01_02/DN_BitueberIEEE802.pdf

[3] Seminar Rechnernetze, Drahtlose Hochleistungskommunikation
http://www.thorsten-karl.de/wlan/

[4] Netzwerkguide, http://www.i-m.de/home/datennetze/ef_qos2.htm

[5] IEEE 802.1, http://grouper.ieee.org/groups/802/1/

[6] wikipedia, http://en.wikipedia.org/wiki/Quality_of_Service

[7] Realtime Traffic Flow Measurement, http://www2.auckland.ac.nz/net/Internet/rtfm/

Kapitel 6

Moderne Konzepte für Sicherheit in Mobilen Netzen

Jonathan Albe

Während kabellose Netzwerke inzwischen für jedermann zugänglich und erschwinglich sind, hinkt die Sicherheit dieser Entwicklung hinterher. Nach der Einführung der WEP–*Verschlüsselung im Standard* IEEE 802.11 *wurde dessen Unsicherheit sehr schnell festgestellt. Mit der zunehmenden Verbreitung von WLAN wächst seitdem auch der Bedarf nach Sicherheit in diesen Netzen. Der Standard* IEEE 802.11i *soll die bestehenden Sicherheitslücken in kabellosen Netzwerken schließen. Das in* IEEE 802.11i *enthaltene* TKIP *Verfahren sowie die Möglichkeit der Authentifizierung durch* IEEE 802.1x *stehen einem Netzbetreiber bereits durch WPA zur Verfügung. Wird der neue Standard ratifiziert, erscheint damit das Verschlüsselungsprotokoll* CCMP *auf dem Markt, welches durch die Verwendung von* AES *endlich einen höheren Sicherheitsstandard verspricht.*

Die Grundlagen der in IEEE 802.11 *und* IEEE 802.11i *verwendeten Protokolle, ihre Stärken und Schwächen sind Inhalt dieser Arbeit. Auch protokollunabhängige Sicherheitsmaßnahmen werden erläutert.*

Inhaltsverzeichnis

6.1 Einleitung 123
6.2 Sicherheitsrisiken und Angriffsszenarien für Netzwerke . . . 123
6.3 Sicherheitsansprüche für Netzwerke 124
6.4 IEEE 802.11 WEP 125
6.4.1 WEP – Funktionsweise 125
6.5 WPA – Zwischenlösung 129
6.6 IEEE 802.11i Neue Sicherheit 129
6.6.1 IEEE 802.11i – TKIP 129
6.6.2 IEEE 802.11i – CCMP 131
6.7 IEEE 802.1x - Authentifizierung und Schlüsselmanagement . 137
6.7.1 IEEE 802.1x - Kommunikationsprotokolle 138
6.7.2 IEEE 802.1x Authentifikation durch RADIUS 139
6.8 Konfigurationsmanagement 139
6.8.1 Mischbetrieb von WEP und TKIP 139
6.8.2 Schlüsselmanagement 140
6.8.3 Abschotten durch Firewall 140
6.9 Maßnahmen zur Erhöhung der Sicherheit in Netzwerken . . 142
6.10 Zusammenfassung 143

6.1 Einleitung

Mit heute verfügbaren Produkten ist es denkbar einfach ein kabelloses Netzwerk (Wireless Network) einzurichten und in Betrieb zu nehmen, so dass zwei oder mehrere Teilnehmer Daten kabellos austauschen können. Der Nutzer muss lediglich einen Zugangspunkt (Access Point) montieren und diesen eventuell an ein bereits vorhandenes Netzwerk oder zum Beispiel einen DSL-Router anschließen. Auf den Stationen muss ein WLAN[1] Adapter installiert werden, was im günstigsten Fall durch Plug & Play eine denkbar einfache Aufgabe ist. Aus Sicht des Benutzers entsteht so ein Netzwerk ohne aufwendiges Verlegen von Kabeln und den damit verbundenen Hindernissen. Der große Vorteil des kabellosen Netzwerks ist zugleich sein größter Nachteil. Daten werden mittels elektromagnetischer Wellen "durch die Luft" übertragen. Die Ausbreitung dieser Wellen kann zwar durch Hindernisse wie Gebäudeteile eingeschränkt werden, ist jedoch generell nicht vorhersagbar. Und auf keinen Fall kann davon ausgegangen werden, dass sich ihre Ausbreitung auf das eigene Haus, das Büro oder ein Firmengelände beschränkt.

Während bei drahtgebundenen Netzwerken ein direkter physischer Zugang zum Kabel gegeben sein muss, können WLAN Übertragungen relativ einfach, durch bloße Nähe und ggf. eine Zusatzantenne abgehört und manipuliert werden. Diese Problematik ist hinreichend bekannt und zog bereits einige Sicherheitsprotokolle und Verfahren nach sich, die jedoch keinesfalls ausreichten um Funknetzwerke zu schützen. Es herrscht nach wie vor ein starker Bedarf an Sicherheitsmechanismen im Zusammenhang mit drahtlosen Netzen.[1]

Diese Seminararbeit betrachtet die aktuellen Entwicklungen und zeigt Vor- bzw z. Nachteile gegenüber vorhergehenden Sicherheitsmechanismen.

6.2 Sicherheitsrisiken und Angriffsszenarien für Netzwerke

Um Sicherheitsalgorithmen und Protokolle beurteilen zu können, ist es notwendig die Möglichkeiten und Methoden eventueller Angreifer zu untersuchen. Im Folgenden werden einige grundlegende Methoden aufgezeigt um in Netzwerke einzudringen oder diese zu manipulieren. Angriffswege die auf *Social Engineering*, also das Ausnutzen, Täuschen oder Erpressen von Menschen abzielen, sind nicht Inhalt dieser Arbeit und sollen an dieser Stelle nicht betrachtet werden. Der Schwerpunkt liegt hier auf den technischen und kryptografischen Methoden [2].

- **Spoofing** bezeichnet das Verbergen der eigenen Identität hinter einer anderen. In ungeschützte Funknetzwerke kann sehr einfach eingebrochen werden, wenn der Angreifer eine vorhandene MAC[2] / IP Adresse ausliest und fortan eigene Daten unter dieser Adresse sendet. Tools wie zum Beispiel `Kismet` [3] erleichtern dieses Vorgehen.

[1] WLAN: Wireless LAN, Funknetzwerk
[2] Medium Access Control

- **Sniffing** beschreibt das passive Mithören der übertragenen Daten. Es kann so auf ungesicherte Daten zugriffen werden. Auch bei einer geschützten Verbindungen ist das passive Zuhören sinnvoll um Informationen zu sammeln, falls die Verschlüsselung bzw. Schutzprotokolle angegriffen werden sollen.
- **Session Hijacking** ist auch als `Man-in-the-middle-Attack` bekannt. Ein Angreifer klinkt sich zwischen zwei Kommunikationspartner, leitet die Daten über einen eigenen Rechner um und kann diese so manipulieren. Solche Angriffe sind z.B. mit gefälschten Statusmeldungen möglich, welche eine Umleitung von Paketen verursachen können.
- **Replay Attacken** versuchen das Netz, oder darin befindliche Programme, durch ein späteres Neusenden von Paketen zu beeinflussen. Hierzu werden Pakete zwischengespeichert und zu einem späteren Zeitpunkt erneut gesendet [4].
- **DoS**[3] **Attacken** sind eher Angriffe auf Ressourcen, als auf Sicherheitsmechanismen. Sie werden zum einen der Vollständigkeit wegen erwähnt, zum anderen werden sie später im Zusammenhang mit Abwehrmechanismen noch einmal betrachtet. Hierbei wird das Ziel systematisch Überlastet, um dessen Verfügbarkeit zu stören.
- **Brute Force Attacken** bezeichnen ein Verfahren, mit dem Kennwörter oder Schlüssel durch systematisches Probieren herausgefunden werden können. Die Dauer dieses Vorgangs hängt vor allem von der Rechenleistung des Angreifers und der Länge des Passwortes ab. Bei einer ungünstigen Passwortvergabe oder Benutzerfehlern kann dieser Prozess deutlich beschleunigt werden. Insbesondere bei der Implementierung und Vergabe von Passwörtern muss diese Technik bedacht werden.

6.3 Sicherheitsansprüche für Netzwerke

Aus der Betrachtung möglicher Bedrohungen und eigener Interessen entstehen einige Ansprüche, welche an Sicherheitsprotokolle zu stellen sind. [5]

- **Verfügbarkeit** bezieht sich im Zusammenhang auf Netzwerkverbindungen in erster Linie auf das Bereitstehen der Verbindung und der Möglichkeit zu kommunizieren. Im Rahmen dieser Arbeit bedeutet Verfügbarkeit nicht die physikalische Verbindung zu einem Netzwerkteilnehmer oder einem Zugangspunkt, sondern vielmehr die Möglichkeit diese auch ansprechen zu können. Ein Zugangspunkt, welcher unter der Last zu vieler Pakete wegen eines DoS-Angriffes oder aufgrund falscher Anweisungen abschaltet, ist eben nicht mehr verfügbar.
- **Vertraulichkeit** verhindert, dass unbefugte Dritte Kenntnis von Daten erlangen. In einem Funknetz müssen hier leistungsfähige und sichere Algorithmen zum Einsatz kommen um die Vertraulichkeit zu gewährleisten, da die Übertragung selbst nicht geschützt werden kann.
- **Integrität** steht für die Authentizität der übertragenen Daten. Der Empfänger muss wissen, ob die gesendeten Daten genau die durch den Sender gesendeten sind oder

[3]Denial of Service: Nichtverfügbarkeit eines Dienstes

ob Störungen bzw. bewusste Veränderung durch Dritte die Daten verfälscht haben. Diese Erkennung ist im Zusammenhang mit Vertraulichkeit besonders wichtig, denn ein Erkennen des Verlustes der Vertraulichkeit ist ebenso wichtig wie diese selbst.

- **Authentizität** ist die Basis für eine Zugangskontrolle und eng mit der Integrität verbunden. Während die Integrität den Schutz der Daten an sich gewährleistet, befasst sich die Authentizität mit dem Schutz von Quelle und Ziel dieser Daten.
- **Zugangskontrolle** schützt Ressourcen und Daten, indem nur berechtigte Benutzer Zugang zu diesen erhalten.

Neben diesen elementaren Sicherheitsansprüchen ergeben sich weitere Punkte, die beim Entwurf von Sicherheitsprotokollen für Funknetzwerke bedacht und berücksichtigt werden müssen. Dazu gehören:

- **Ein angemessener Administrationsaufwand**: Weitere Informationen zum Administrationsaufwand und Konfigurationsmanagement enthält Kapitel 6.8
- **Der Rechenaufwand der Zugangspunkte**: ist oft beschränkt, wird jedoch für die Verschlüsselung der Nutzdaten benötigt. Im Allgemeinen muss davon ausgegangen werden, dass grundsätzlich neue Verschlüsselungsverfahren auch neue Hardware benötigen.
- **Quality of Service in Wireless Networks**: Eine Steigerung der Sicherheit ist im Allgemeinen mit mehr Aufwand und einer Minderung der Qualität des Netzes, z.B. der Reaktionsgeschwindigkeit oder der Bandbreite, verbunden. Dieser Aspekt ist ebenfalls zu berücksichtigen.

6.4 IEEE 802.11 WEP

Den meisten Funknetzen liegt der IEEE Standard 802.11 zugrunde. Insbesondere das in 802.11 enthaltene WEP (Wireless Equivalent Privacy) Protokoll wurde jedoch schon kurz nach seiner Veröffentlichung als zu schwach erkannt und gebrochen [6].

Aber auch wenn WEP nicht als sicher gilt, ist es dennoch weit verbreitet und Grundstein für einige nachfolgende Sicherheitsprotokolle. Aus diesem Grund werden WEP und seine Schwachstellen nachfolgend betrachtet.

6.4.1 WEP – Funktionsweise

WEP ist ein symmetrisches Verschlüsselungsverfahren, das auf der RC4 Stromverschlüsselung von RSA beruht.

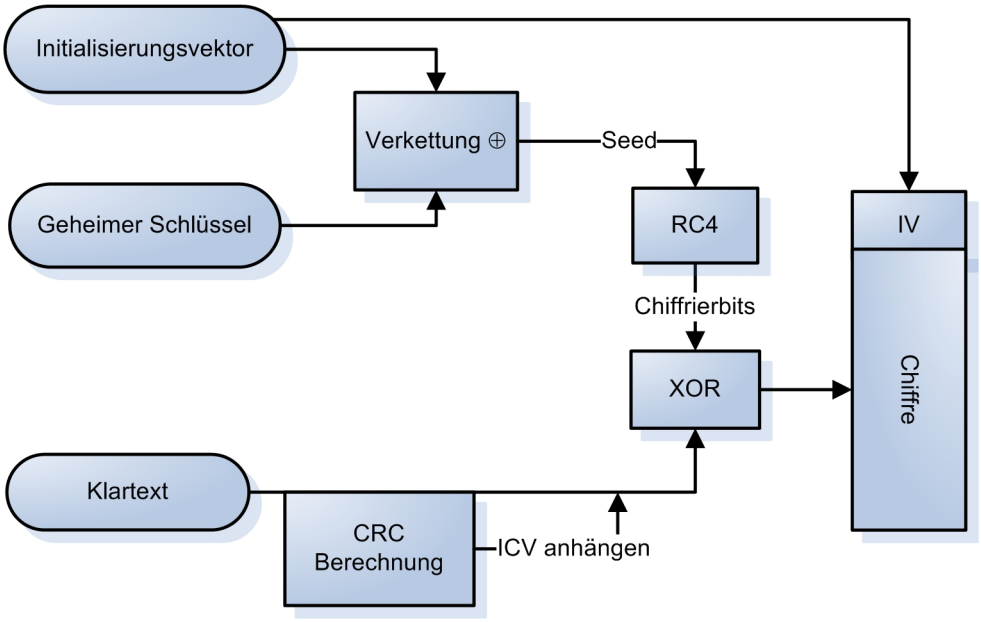

Abbildung 6.1: WEP – Verschlüsselung

WEP Verschlüsselung

Abbildung 6.1 zeigt die wesentlichen Elemente dieses Vorgangs[7, 24].

Am Anfang des Prozesses liegen beim Sender der Klartext so wie der geheime Schlüssel (Shared Key) vor. Je nach WEP Version hat der Schlüssel eine Länge von 40 oder 104 Bit. Dieser Schlüssel bildet, zusammen mit dem 24 Bit langen Initialisierungsvektor eine Zustandsreihe, den Seed. Die Verknüpfung zwischen Schlüssel und einem Vektor, der sich ausgehend vom Initialisierungsvektor verändert, soll verhindern, dass gleiche Nutzdaten auf die gleiche Weise verschlüsselt werden. Die Erzeugung des Initialisierungsvektors obliegt der jeweils verwendeten Hardware/ Firmware.

Der Seed wird von einem Pseudo Random Nummer Generator (PRNG) genutzt um einen Strom von Chiffrierbits zu verwenden. Dieser PRNG verwendet zur Erzeugung des Bitstroms den RC4 Algorithmus der Firma RSA [8]. Der so entstandene Bitstrom wird zur eigentlichen Verschlüsselung verwendet und mit dem “Klartext“, also den zu versendenden Nutzdaten, XOR verknüpft. Vor dieser Verknüpfung, der eigentlichen Verschlüsselung, wird eine CRC Berechnung durchgeführt und der ICV (Integrity Check Value) an den Klartext angehängt. Das Ergebnis der Verbindung zwischen Klartext und den Chiffrierbits sowie der Initialisierungsvektor werden anschließend übertragen.

WEP Entschlüsselung

Der Empfänger nimmt die übertragenen Nutzdaten so wie den Initialisierungsvektor entgegen. Mit Hilfe des übertragenen Initialisierungsvektors und dem geheimen Schlüssel, welcher beim Empfänger ebenfalls bekannt sein muss, ist der Empfänger in der Lage den gleichen Seed und somit die identische Folge von Chiffrierbits zu bilden. Durch XOR

Verknüpfung kann der Empfänger so den Klartext erlangen. Mit Hilfe des angehängten Integrity Check Value können die Nutzdaten auf Fehler und Veränderung geprüft werden. Abbildung 6.2 zeigt diesen Vorgang.

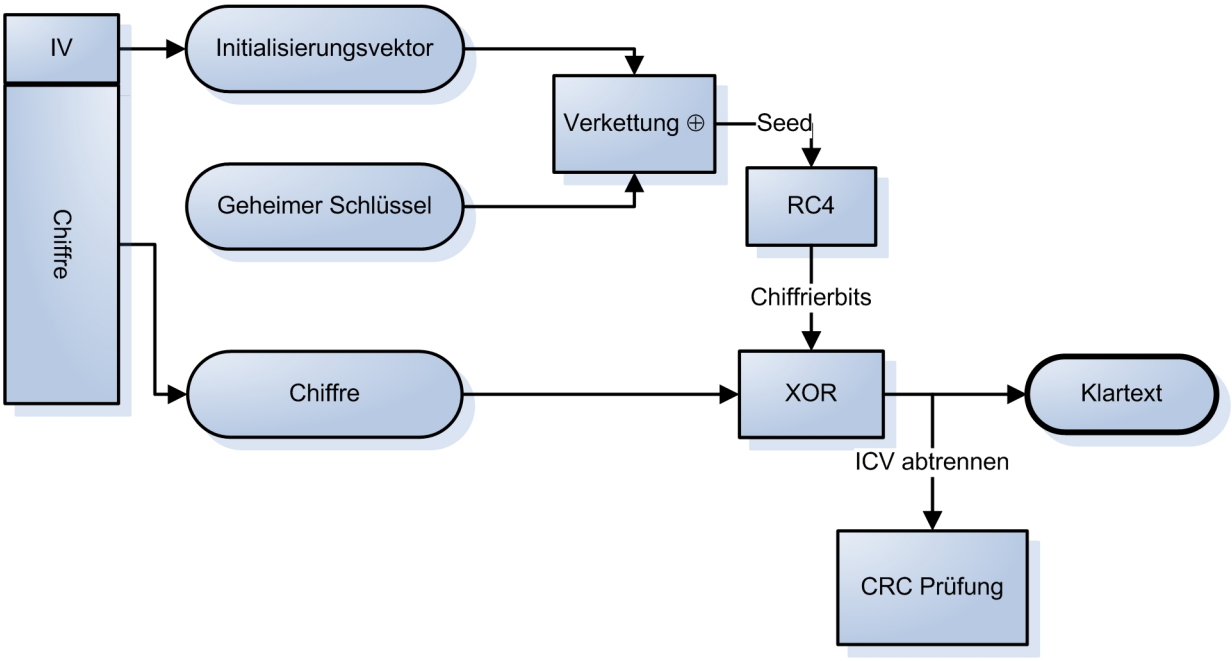

Abbildung 6.2: WEP – Entschlüsselung

WEP – Designfehler und Schwächen

Bereits kurz nach der Einführung von WEP wurden dessen Schwächen offensichtlich [6] und dieses Protokoll gebrochen. Dabei wurde gezeigt, dass es auch in der Praxis relativ einfach ist WEP passiv zu brechen [9].

Im Nachfolgenden sollen die größten Fehler von WEP gezeigt werden. Für einen möglichen Angreifer ist hier die Seite des Senders interessant. Abbildung 6.3 markiert anhand des Verschlüsselungsschemas die wesentlichen Schwachstellen.

1. **Fehlendes Schlüsselmanagement:** WEP besitzt keinerlei Schlüsselmanagement. Der geheime Schlüssel, welcher für die verschlüsselte Kommunikation genutzt wird, findet auch bei der Authentifizierung mittels Challenge-Response Verfahren seine Anwendung. Wer also die Authentifizierung bricht, kann den verschlüsselten Funkverkehr abhören — und umgekehrt. Weiterhin wird für die gesamte Kommunikation nur ein Schlüssel verwendet. Eine Tatsache die WEP in Verbindung mit einem ohnehin schon kurzen Schlüssel extrem angreifbar macht.
2. **Erzeugung des Initialisierungsvektors:** Bestimmte schwache Initialisierungsvektoren machen die Verschlüsselung weiter angreifbar, da sie wiederum zu schwachen und statistisch auswertbaren Schlüssel-Chiffrierfolgen führen [11].
3. **Zu kurzer Initialisierungsvektor:** Die Länge von 24 Bit erlaubt in Verbindung mit einem einzigen festen Schlüssel eine maximale Kombination von 2^{24} verschiedenen

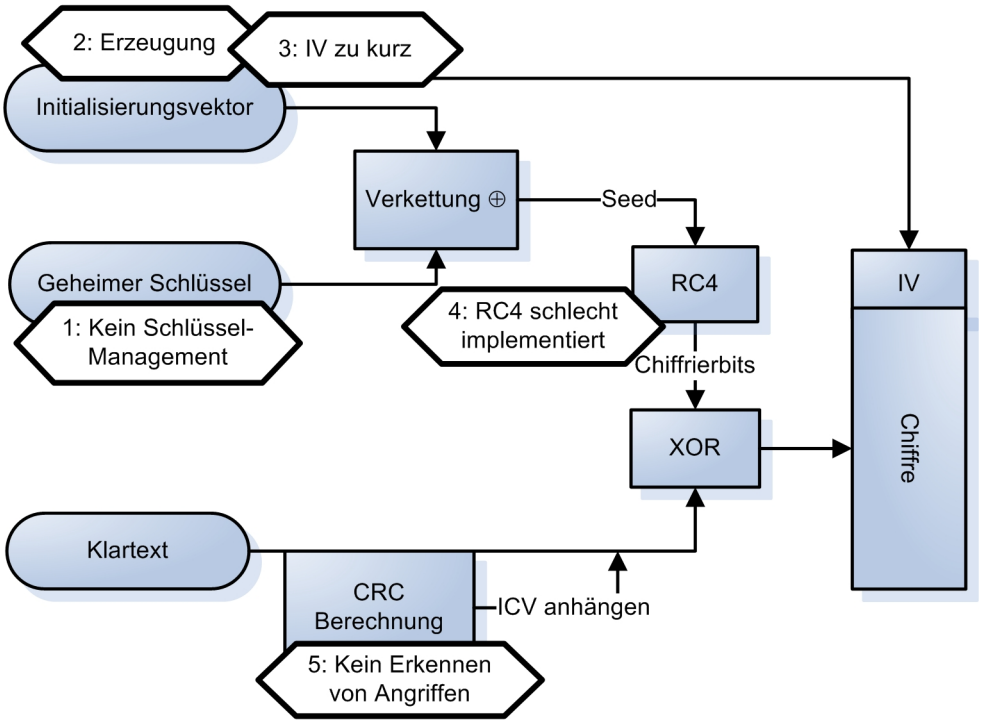

Abbildung 6.3: WEP – Schwachstellen

Schlüsseln. Nach maximal 2^{24} Schlüsseln muss sich also der Initialisierungsvektor wiederholen. Eine einfache Rechnung zeigt, dass bei Paketgrößen von 1000 Byte pro Paket und einer Übertragung von 11 MBit, nach ca. 3,5 Stunden 2^{24} Pakete übertragen wurden. Wenn ein Netzwerk nicht unter maximaler Last läuft vergrößert sich dieser Zeitraum, jedoch kann ihn eine ungünstige Weiterschaltung des Initialisierungsvektors wieder verkürzen. Es ist also offensichtlich, dass wenige Stunden Aufzeichnung ausreichen um verschiedene Pakete mit gleichem Initialisierungsvektor zu empfangen. Diese Kombination ist mit mathematischen und statistischen Methoden angreifbar[4]. Befinden sich mehrere Access Points und Teilnehmer in einem Netz sinkt die Zeit, welche für ein passives Brechen von WEP benötigt wird, mit dem steigenden Funkverkehr weiter.

4. **Schlechte Implementierung von RC4:** In der verwendeten Implementierung von RC4 treten regelmäßig schwache Folgen von Chiffrierbits auf, welche besonders leicht gebrochen werden können[11]. Viele Tools zum Brechen von WEP, z. B. AirSnort [12], nutzen diese Schwachstelle aus.

5. **Unzureichender Schutz der Daten:** Der Schutz von übertragenen Daten wird in WEP durch einen linearen Algorithmus bereitgestellt. Diese Prüfsumme kann selbst ohne Kenntnis des WEP-Schlüssels manipuliert werden. Besitzt der Angreifer erst den WEP-Schlüssel, stellt der Integrity Check Value kein Hindernis mehr dar.

[4] [7] demonstriert einen Angriff auf WEP

6.5 WPA – Zwischenlösung

Der unzureichende Schutz durch WEP wurde schnell erkannt und publiziert [6, 9]. Ungeachtet dessen herrscht ein massiver Bedarf an Sicherheit in drahtlosen Netzwerken und das Erscheinen neuer, verbesserter Protokolle verzögert sich weiterhin [13]. Diese Tatsache veranlasste mehrere Hersteller von WLAN Produkten zum Handeln. Die Wi-Fi[5] Allianz [14], ein Zusammenschluss vieler großer Hersteller von WLAN Produkten, vertreibt aus diesem Grund eine Zwischenlösung Namens WPA [15]. WPA steht für Wi-Fi Protected Access und kann auf Hardware mit Wi-Fi Zertifikat eingesetzt werden. WPA nutzt also bestehende Hardware und versucht die Sicherheitslücken von WEP zu schließen. Wesentlicher Bestandteil von WPA sind bereits fertig gestellte Teile des Standards 802.11i, vor allem das Protokoll TKIP auf welches im Kapitel 6.6.1 näher eingegangen wird. Für große Netzwerke kann WPA auf weitergehende Authentifizierungsmethoden zurückgreifen, welche im Kapitel 6.7 beschrieben werden. Wichtige Merkmale von WPA sind: Abwärtskompatibilität zu bereits vertriebenen Produkten auf dem Markt, die Möglichkeit WPA durch ein Softwareupdate auf vorhandene Produkte zu installieren und Aufwärtskompatibilität zum Standard 802.11i der voraussichtlich Mitte 2004 ratifiziert wird [16].

6.6 IEEE 802.11i Neue Sicherheit

Die Sicherheit in kabellosen Netzwerken soll mit der Einführung des neuen Standards 802.11i deutlich angehoben werden. Eine Anforderung an einen solchen neuen Standard ist die Abwärtskompatibilität, welche ein Betreiben auf bereits vorhandener Hardware möglich macht und letztendlich eine schrittweise Einführung neuer Sicherheitsmechanismen ermöglicht. Dies ist in so fern problematisch, als die Einführung neuer Verschlüsselungsalgorithmen im Allgemeinen auch neue Hardware benötigt. 802.11i enthält darum zwei getrennte Sicherheitsprotokolle. TKIP für die Verwendung auf bereits vorhandener Hardware um eine Abwärtskompatibilität zu gewährleisten und CCMP als völlig neues, auf AES basierendes Verschlüsselungsprotokoll. Des weiteren stellt 802.11i die Möglichkeit zur Verfügung, Methoden zum Schlüsselmanagement und zur Authentifizierung zu nutzen. Diese werden im Kapitel 6.7 erläutert.

6.6.1 IEEE 802.11i – TKIP

TKIP steht für "Temporal Key Integrity Protocol" und ist die Zwischenlösung für Abwärtskompatibilität in 802.11i. TKIP kann durch Treiberupdates in bereits vorhandene Hardware implementiert werden. Aus eben diesen Kompatibilitätsgründen ist der RC4-Algorithmus weiterhin der zentrale Bestandteil der Verschlüsselung. Wesentlicher Unterschied zu WEP ist die Verschlüsselung der Nachrichten durch einen temporären Schlüssel [17]. Abbildung 6.4 zeigt die grobe Struktur der TKIP Verschlüsselung.

[5] Wireless Fidelity

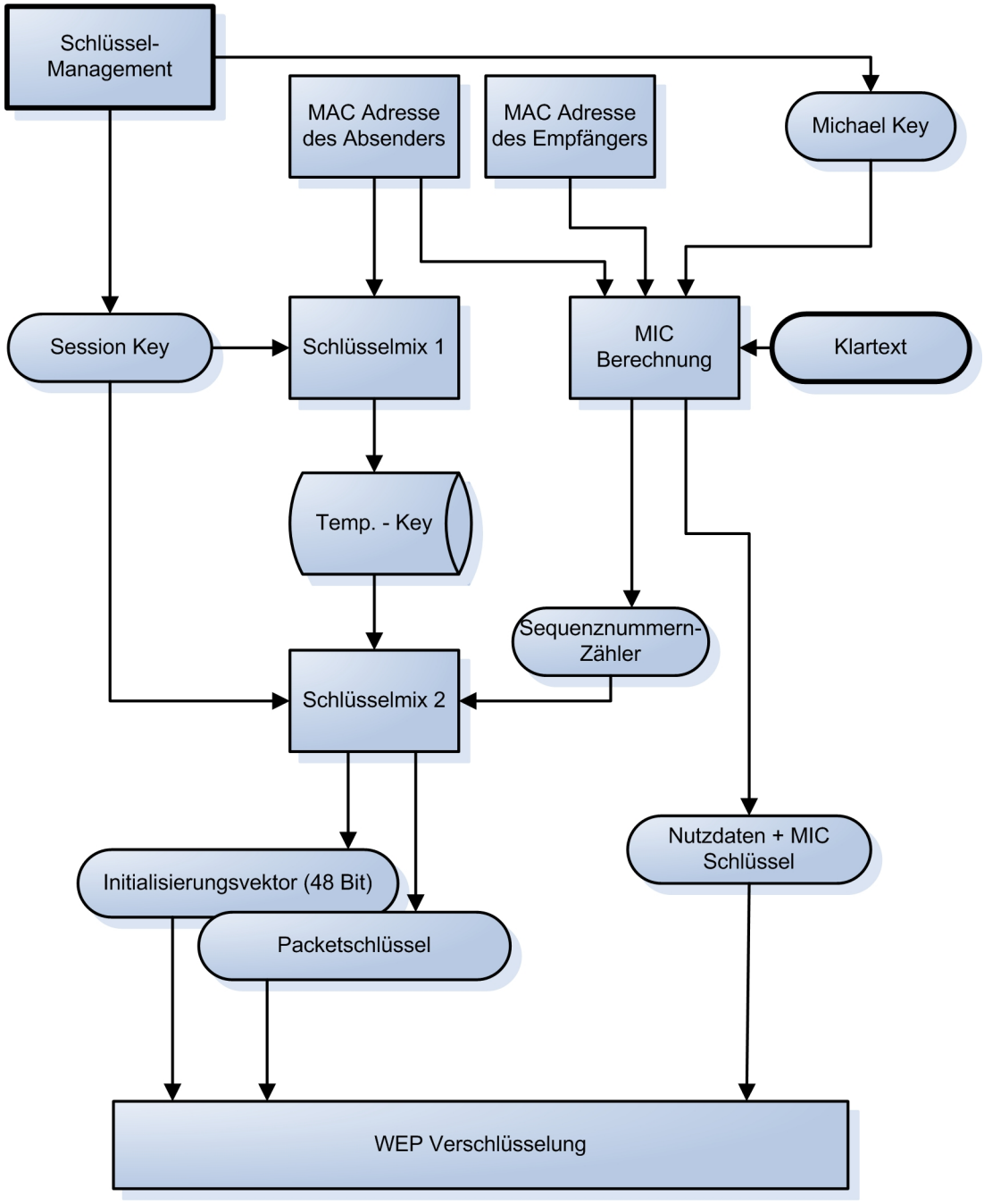

Abbildung 6.4: TKIP – Verschlüsselung

Ausgangspunkt der TKIP Verschlüsselung ist ein temporärer Sitzungsschlüssel (Session Key), welcher für die laufende Sitzung verwendet wird. Dieser Schlüssel wird vom Schlüsselmanagement bereitgestellt. Auf das Schlüsselmanagement wird in Kapitel 6.7 näher eingegangen. Mit den zwei verschiedenen Schlüsselmixverfahren und den Parametern, welche darauf Einfluss nehmen, sollen die Schwächen des statischen Schlüssels und des Initialisierungsvektors umgangen werden.

Schlüsselmix Eins erzeugt einen Temporärschlüssel aus dem Session Key und der MAC Adresse des Absenders. Das Einbinden der eigenen MAC Adresse verhindert, dass verschiedene Sender identische Daten auf die gleiche Weise verschlüsseln. Das Ergebnis der relativ aufwendigen ersten Phase kann zwischengespeichert werden.

Die MIC Berechnung schützt die Integrität der Nachrichten. MIC steht für Message Integrity Code und wird mit der MAC Quell- und Zieladresse, einem MIC Schlüssel und den eigentlichen Nutzdaten berechnet. Der MIC schließt die gesamte Nachricht, und nicht einzelne Pakete, ein.

Zum Schutz vor Replay Attacken wird eine Sequenznummer in die zweite Schlüsselmixprozedur eingebunden. Das Ergebnis der zweiten Schlüsselerzeugung wird dem RC4 Algorithmus zugeführt. Die durch den MIC geschützten Nutzdaten werden anschließend durch die Folge von Chiffrierbits aus dem RC4 Algorithmus verschlüsselt, ganz wie es bei WEP der Fall ist.

Gegenmaßnahmen

Der Schutz der Integrität ist aufgrund der geforderten Kompatibilität nur eingeschränkt möglich. Die Berücksichtigung der Rechenleistung von Zugangspunkten verursacht eine kurze und somit angreifbare Prüfsumme. Um dennoch Sicherheit zu gewährleisten, wurden im TKIP Gegenmaßnahmen implementiert, welche die Verbindung schützen sollen. TKIP registriert auf der Empfängerseite Eingriffe in geschützte Nachrichten und schreibt diese in ein LOG. Stellt TKIP innerhalb von 60 Sekunden mehrere Verletzungen fest, so wird das Netz für 60 Sekunden blockiert und die Sitzungsschlüssel so wie eventuelle Authentifikationen müssen erneuert werden.

Durch die gezielte Abschaltung ist es theoretisch möglich, ein durch TKIP geschütztes Netz mit wenigen geänderten Nachrichten lahm zu legen. Diese Schwäche von TKIP wird durch die Tatsache relativiert, dass ein WLAN mit deutlich weniger Aufwand und wesentlich mehr Paketen ebenfalls lahm gelegt werden kann.

6.6.2 IEEE 802.11i – CCMP

Während TKIP aus Kompatibilitätsgründen einen Kompromiss zwischen alter Hardware und neuer Sicherheit schließt, ist CCMP ein gänzlich neu designtes Sicherheitsprotokoll. Es beruht auf dem Advanced Encryption Standard (AES). CCMP steht für CTR/ CBC-MAC. Dabei bezeichnet CRT (Counter) die Betriebsart von AES und CBC-MAC Cipher Block Chainig – Message Authentication Code. Die Bedeutung dieser Abkürzungen wird in den nachfolgenden Absätzen deutlich.

AES Grundlagen

AES ist ein Blockchiffre. Im Vergleich zum Bit für Bit arbeitenden RC4 Algorithmus in WEP, werden die Daten unter CCMP blockweise verschlüsselt. Aus diesem Grund kann CCMP auf Paketebene arbeiten, das Aufteilen (Fragmentieren) von Nachrichten kann auf höherliegenden Ebenen stattfinden. Während AES allgemein in der Lage ist Blöcke variabler Größe zu verschlüsseln wurde die Blockgröße in CCMP auf 128 Bit festgelegt.

AES kann in verschiedenen Betriebsmodi verwendet werden. Im Modus ECB (Electronic Codebook) werden die einzelnen Blöcke direkt durch AES verschlüsselt. Und auch wenn der ECB - Modus in CCMP nicht verwendet wird, so soll er kurz aufgezeigt werden um die Wahl des nachfolgenden Modus zu begründen. Abbildung 6.5 zeigt AES Verschlüsselung im ECB Modus.

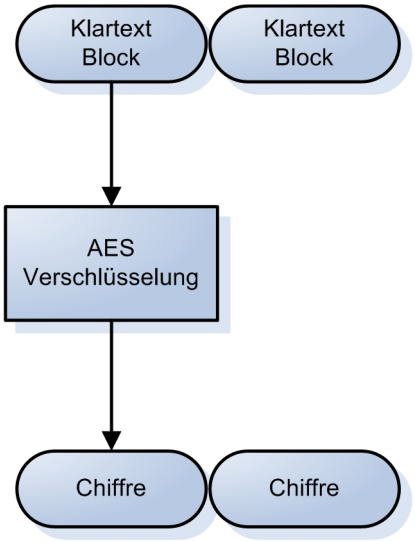

Abbildung 6.5: AES – Electronic Codebook

Die alternative Variante zu ECB ist der “Counter Modus“ (CTR). Hier wird statt der Nachricht ein Zählwert verschlüsselt und das Resultat mir den Nutzdaten XOR verknüpft. Abbildung 6.6 zeigt AES im Counter Modus [18]. Folgende Vorteile ergeben sich durch die Verwendung des Counter Modus im Vergleich zum ECB Modus:

- Durch die XOR Verknüpfung am Ende spielt es keine Rolle ob das letzte Paket genau 128 Bit groß ist, ein eventuelles Hinzufügen von Füllbits entfällt.
- Gleiche Klartextblöcke ergeben verschiedene Chiffreblöcke. In Netzwerken wo sich bestimmte Muster, wie z. B. TCP-Header oder bestimmte Steuerungssignale wiederholen wäre dies ein Nachteil.
- Die in der Grafik eingerahmten Schritte, Berechnen und Verschlüsseln des Zählers, können parallel ausgeführt werden, da sie nach dem Initialisierungswert nicht mehr vom Inhalt der Pakete abhängen. Wie in der Grafik 6.6 ersichtlich, ist der Codier- und Decodiervorgang exakt derselbe, es ist keine unterschiedliche Hard- oder Software nötig.

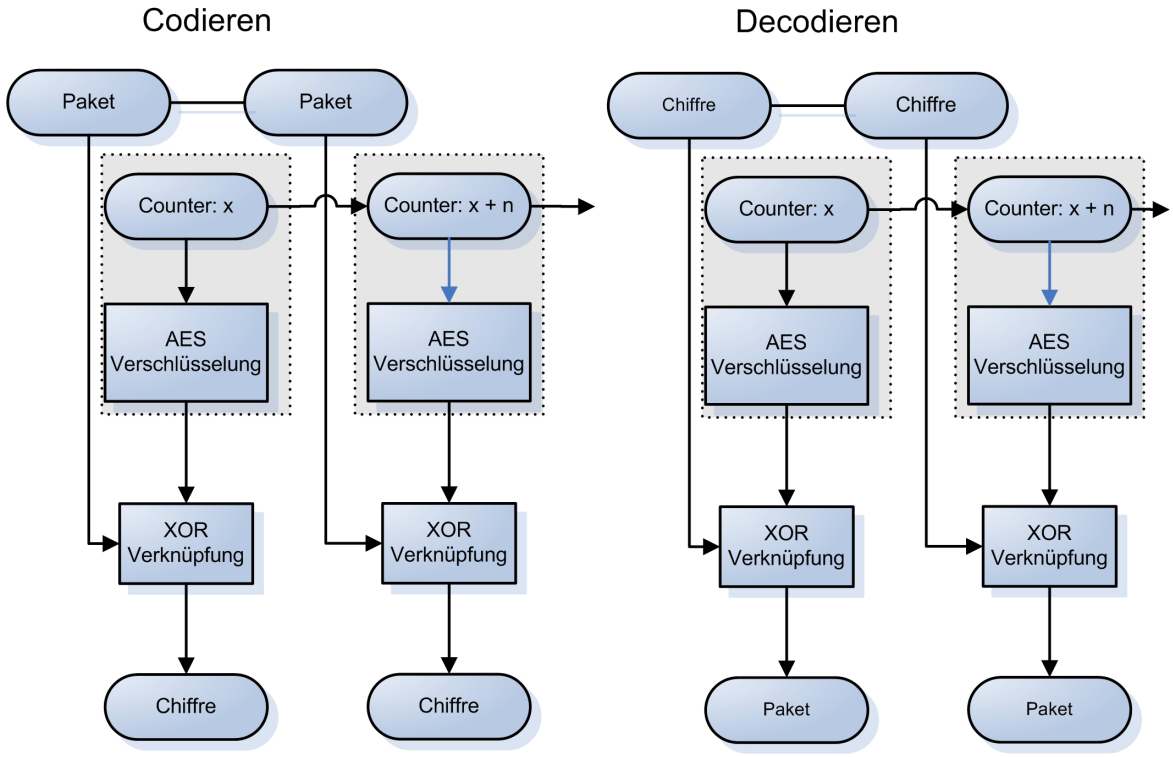

Abbildung 6.6: AES – Counter

CCMP – Verschlüsselung

Abbildung 6.7 zeigt den Verschlüsselungsvorgang im CCMP Protokoll (vereinfacht) [19]. Die Verschlüsselung geschieht hier auf Paketebene. MDPU am Anfang und am Ende bezeichnet die ein und ausgehenden Pakete, die MAC Protocol Data Unit. Der temporäre Sitzungsschlüssel wird auch hier durch das Schlüsselmanagement bereitgestellt (Siehe Kapitel 6.7). Aus der Sequenznummer und der Senderadresse wird ein Nonce Wert konstruiert. Dieser Nonce Wert wird für die Initialisierung der Verschlüsselung und der MIC[6] Berechnung verwendet, um eine wiederholte Verwendung des Sitzungsschlüssels zu verhindern.

Der CCMP Header entspricht in etwa der Übertragung des Initialisierungsvektors bei WEP und TKIP, es werden für die Entschlüsselung relevante Daten übertragen. Die Weitergabe des Message Integritiy Code (MIC), also einer Art Prüfsumme, ähnelt der Weitergabe bei WEP oder TKIP. Der MIC wird an die unveränderten Nutzdaten angehängt und anschließend mit diesen verschlüsselt. Die Berechnung des MIC unterscheidet sich jedoch grundlegend von den bisherigen.

CCMP – Integritätsschutz

Für die Generierung des Message Integrity Code wird die vorhandene AES Verschlüsselung verwendet[19]. Die Nachricht wird in Blöcke aufgeteilt welche nacheinander durch AES

[6]Message Integrity Code: Integritätsschutz der Nachricht

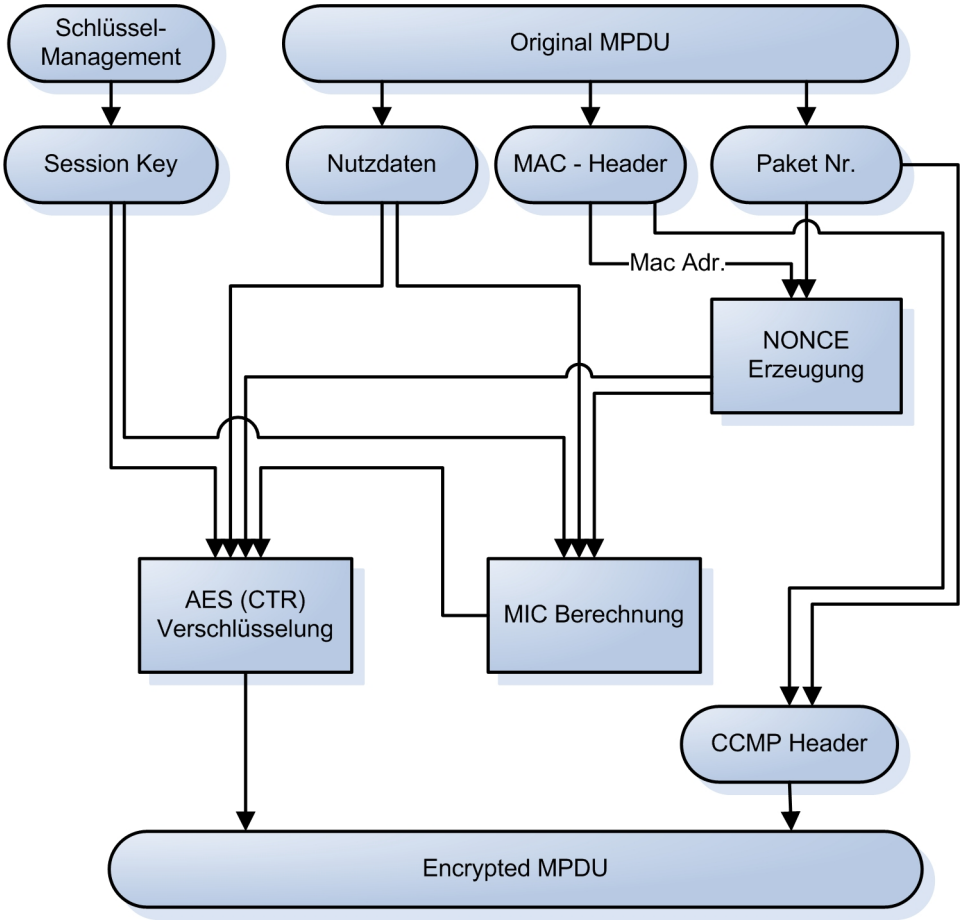

Abbildung 6.7: CCMP – Verschlüsselung

verschlüsselt werden wie Abbildung 6.8 verdeutlicht.

Dieses AES-Verschlüsselungsverfahren wird als CBC-MAC bezeichnet. Neben der Nachricht werden der Session Key sowie der Nonce Wert — und somit die Absenderadresse und Paketnummer — in den MIC eingebunden. Diese Zusatzinformationen garantieren einen eindeutigen MIC ohne Wiederholungen, welcher durch den Sitzungsschlüssel geschützt ist. Diese Berechnung ist eine Einwegfunktion und muss beim Empfänger mit den gegebenen Ausgangswerten und der entschlüsselten Nachricht wiederholt werden. Stimmt der übermittelte und der neu berechnete MIC überein, so wurde die Nachricht nicht verändert.

CCMP Entschlüsselung

Abbildung 6.9 zeigt den CCMP Entschlüsselungsvorgang. Der erste Schritt ist die Entnahme der Paketnummer aus dem übertragenen CCMP Paket. Hier wird, wie in normalem Netzwerkbetrieb, auch die Reihenfolge der Pakete geprüft. Sollte ein Angreifer versuchen Pakete durch einen Replay Angriff in anderer Reihenfolge einzuspielen, so wird dies bei der MIC Überprüfung auffallen, da die Paketnummer auch hier Einfluss nimmt. Wird bei der Prüfung der Paketnummer ein Fehler erkannt, so wird das Paket verworfen. Im nächsten Schritt werden die Absenderadresse und Paketnummer aus dem empfangenen

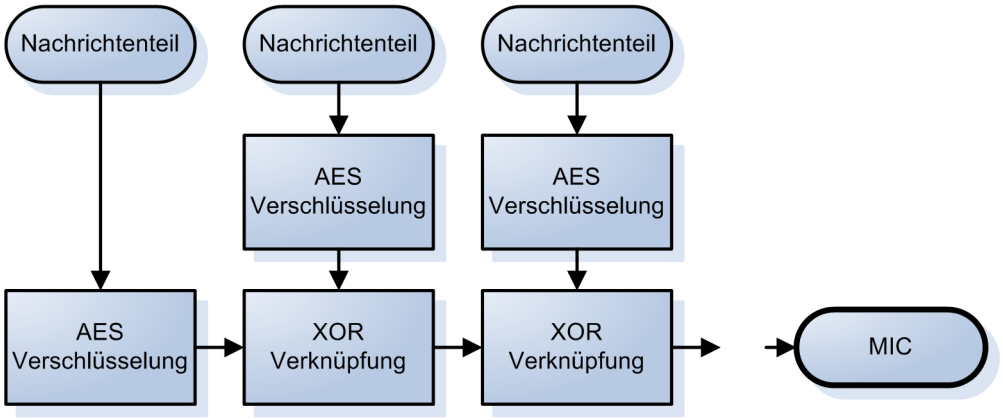

Abbildung 6.8: CCMP – MIC Berechnung

Paket entnommen und zur Konstruktion des Nonce Wertes verwendet. Mit identischer Paketnummer und Absenderadresse entsteht der identische Nonce Wert, wie ihn auch der Absender beim Verschlüsseln verwendet hat. Mit Hilfe des Nonce Wertes, dem Session Key und den Daten des empfangenen Paketes kann die Entschlüsslung durch AES erfolgen. Wie Abbildung 6.6 zeigt ist dies der gleiche Vorgang wie beim Verschlüsseln. Stimmen Session Key und Nonce Wert überein, so verschlüsselt AES die gleichen Counter und erzeugt ein Resultat welches durch XOR Verknüpfung die Originaldaten zurückliefert.

Nach dem Entschlüsseln liegt das Original MPDU Paket und der übertragene Message Integrity Code vor. Wie oben beschrieben wird abschließend ein neuer MIC berechnet um die übertragenen Daten zu verifizieren. Wird hier ein Fehler festgestellt so wird das empfangene Paket verworfen. Aufgrund der deutlich stärkeren Verschlüsselung des MIC Codes entfallen hier Gegenmaßnahmen, wie sie bei TKIP vorliegen.

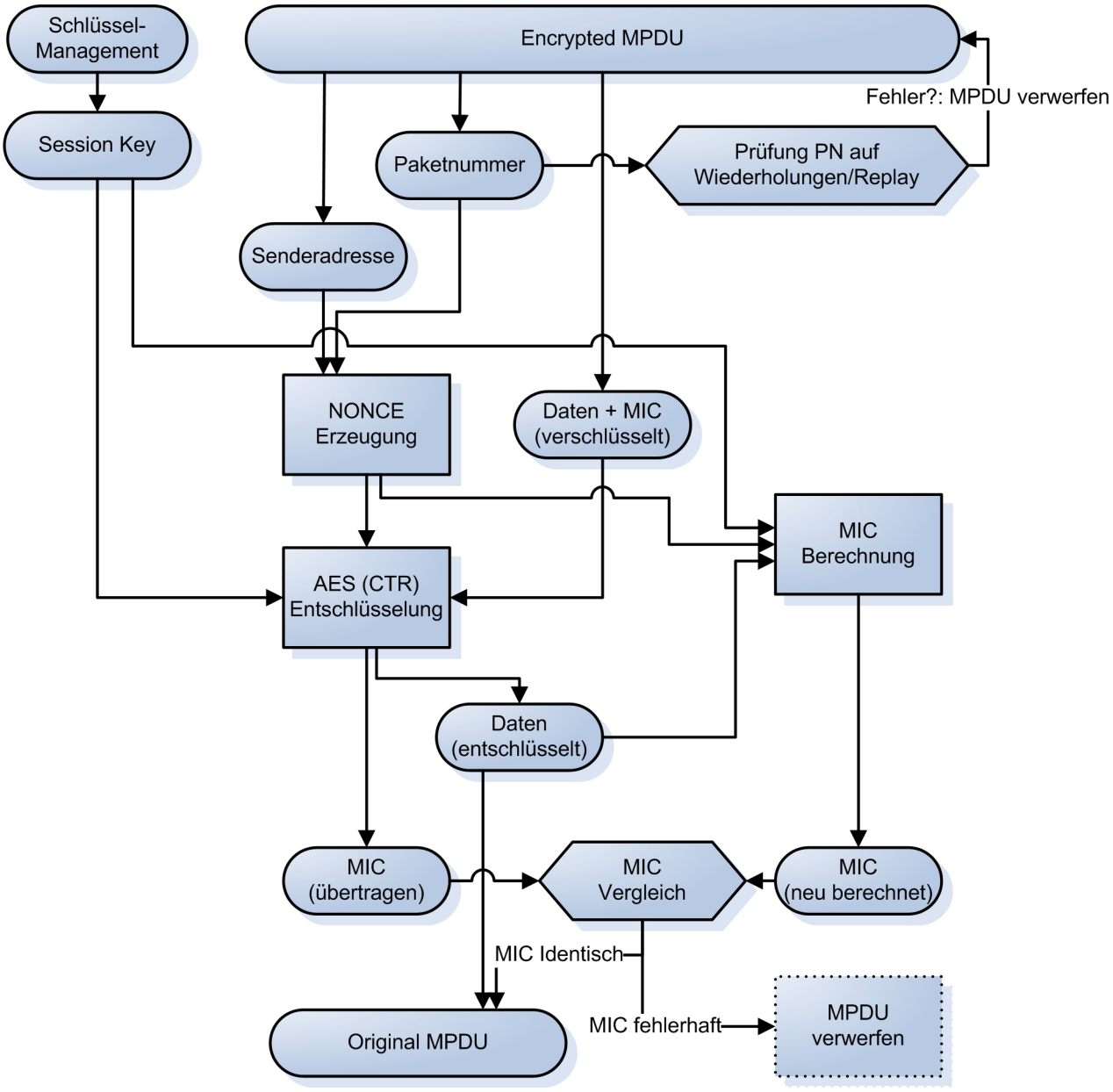

Abbildung 6.9: CCMP - Entschlüsselung und MIC Prüfung

6.7 IEEE 802.1x - Authentifizierung und Schlüsselmanagement

Sowohl TKIP als auch CCMP benötigen einen temporären Schlüssel, den Session Key (Sitzungsschlüssel). Dieser Schlüssel wird vom Schlüsselmanagement durch einen Hauptschlüssel gebildet. Dieser gemeinsame Hauptschlüssel gelangt bei WEP durch Eingabe an den Geräten zu den Kommunikationsteilnehmern. Dies ist für kleine Netze durchaus sinnvoll und am einfachsten zu realisieren. Mit einer wachsenden Teilnehmerzahl ist diese Lösung jedoch nicht mehr praktikabel (Mehr zum Konfigurationsmanagement siehe auch Kapitel 6.8). Aus diesem Grund besteht in IEEE 802.11i die Möglichkeit das IEEE 802.1x Protokoll für die Authentifizierung und Schlüsselverteilung zu verwenden [15].

IEEE 802.1x definiert eine allgemeine, erweiterbare und auf Ports basierende Zugangskontrolle, welche für Funk- und Kabelnetzwerke gleichermaßen gültig ist. 802.1x unterscheidet zwischen zwei Verbindungsarten. Kontrollierte Ports gestatten dem angeschlossenen Nutzer mit allen angeschlossenen Teilnehmern zu kommunizieren. Ein unkontrollierter Port ist hingegen auf bestimmte Adressen und Bereiche eingeschränkt [20]. Ein unangemeldeter Benutzer kann über einen unkontrollierten Port Zugriff auf einen Authentifizierungsserver erhalten, sich authentifizieren und anschließend einen kontrollierten Port verwenden. Abbildung 6.10 zeigt einen möglichen Ablauf des Verfahrens mit einem Radius Server. Die mobile Station sendet eine Anfrage an den Zugangspunkt(Access Point). Er über-

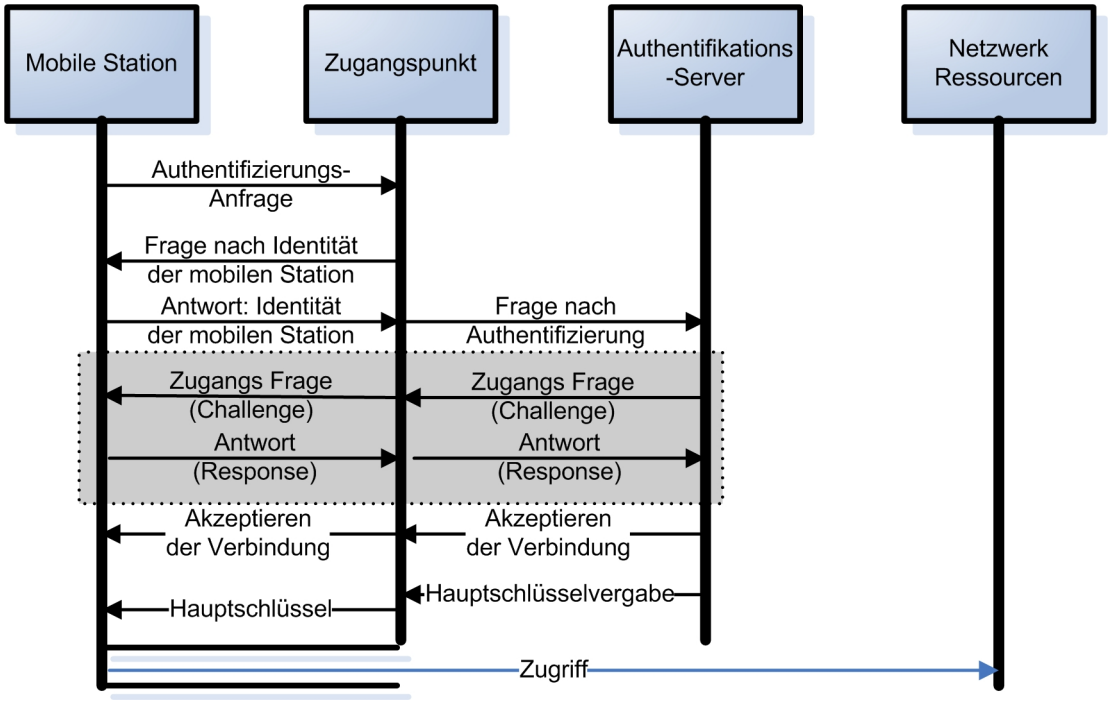

Abbildung 6.10: 802.1x Authentifizierung

nimmt die Rolle des Authentifizierers und vermittelt zwischen der mobilen Station und dem Authentifikationsserver. Der Zugangspunkt leitet die Authentifizierungsanfrage an den Authentifikationsserver weiter. Dieser vergleicht die Anfrage mit den ihm bekannten Benutzerprofilen und gibt (in Abbildung 6.10 eingerahmt) über den Zugangspunkt eine

Challenge – Response Frage an die mobile Station zurück. Akzeptiert der Authentifikationsserver die Antwort der mobilen Station, so informiert er den Zugangspunkt, so dass dieser den entsprechenden Port für die mobile Station öffnen kann und verteilt ggf. die Hauptschlüssel an beide Kommunikationspartner.

6.7.1 IEEE 802.1x - Kommunikationsprotokolle

IEEE 802.1x spezifiziert die Benutzung des Extensible Authentication Protocol[7] (EAP). EAP ist eine Erweiterung des Point to Point Protocol (PPP). EAP kapselt die benötigten Nachrichten und erlaubt es so verschiedene Protokolle für den Authentifizierungsvorgang zu verwenden. Einige mögliche, verschieden starke Protokolle für die Kommunikation bei der Authentifizierung sind [21]:

- **EAP-MD5**
  Hier werden die übertragenen Daten durch einen MD5-Hash Algorithmus codiert. EAP-MD5 authentifiziert Sender und Empfänger nicht! EAP-MD5 verwaltet statische WEP Schlüssel und besitzt kein dynamisches Schlüsselmanagement; es kann also nur zur Identifikation eines Benutzers dienen.
- **EAP TLS**
  Das EAP Transport Layer Security Protocol ist im Wesentlichen eine Kombination zwischen EAP und SSL. Es benötigt eine Authentifizierung auf Basis von Zertifikaten. EAP TLS ermöglicht neben dem benutzerbasiertem auch ein dynamisches Schlüsselmanagement.
- **EAP TTLS**
  EAP - Tunneled Transport Layer Security Protocol ist eine Erweiterung von EAP TLS. Hier benötigt nur der Authentifikationsserver ein Zertifikat. Mit dessen Hilfe wird ein Tunnel zwischen Authentifikationsserver und der mobilen Station aufgebaut, durch welchen die Authentifikation stattfindet. Der Benutzer benötigt hier kein eigenes Zertifikat. Aufgrund dessen ist dieses Protokoll für eine Authentifizierung an Hot Spots geeignet.
- **Cisco Lightweight Authentication Extension Protocol (LEAP)**
  ist eine Lösung der Firma Cisco. Hier authentifizieren sich die mobile Station und der Zugangspunkt gegenseitig. Das dynamische Schlüsselmanagement nutzt eine eigene WEP-Key-Hash Funktion.
- **PEAP Protected EAP**
  erweitert EAP, um Probleme bei der Verwaltung der Zertifikate zu lösen. PEAP verwendet ebenfalls einen TLS Tunnel für die Authentifizierung. Die Verwendung von Upper Layer Protokollen für die Authentifizierung ermöglicht die Einbindung weiterer Software für die Authentifizierung.

Die Möglichkeit verschiedene Protokolle für die Authentifizierung zu verwenden, erlaubt es verschiedene Authentifizierungsverfahren zu implementieren und 802.1x dem eigenen

[7] Erweitertes Authentifikationsprotokoll

Bedarf und der benötigten Sicherheit anzupassen. So ermöglicht EAP-TTLS eine Authentifizierung auf Passwortbasis, während EAP-TLS auf Zertifikaten mit einer Public Key Infrastruktur beruht [22]. Auch eine Authentifizierung durch Smart Cards ist realisierbar [23].

6.7.2 IEEE 802.1x Authentifikation durch RADIUS

Ein verbreiteter Authentifizierungsmechanismus für 802.1x ist die RADIUS – Authentifizierung. RADIUS steht für Remote Authentication Dial–in User Service. RADIUS stellt eine benutzerbezogene Zugangsbeschränkung zur Verfügung [7]. Ein Radius Server ist eine zentrale Instanz, welche in einem Netzwerk Benutzer verwaltet. Die Hauptaufgaben dabei sind Authentifizierung, Autorisierung und Accounting. Neben der Prüfung wer im Netz was tun darf, besteht die Möglichkeit die Nutzung von Ressourcen durch Benutzer zu erfassen um zum Beispiel Dienste abzurechnen.

Die Benutzerverwaltung durch RADIUS ist keine neue Idee für drahtlose Netze. Durch 802.1x kann sie jedoch für Funknetze eingesetzt werden. RADIUS kann mit mehreren Server verwendet werden. Diese Server können in einem Verbund zusammen agieren und Anfragen an jeweils andere RADIUS Server weiterleiten.

6.8 Konfigurationsmanagement

Im Zusammenhang mit der Sicherheit in WLAN, muss das Management der Konfiguration angesprochen werden. In nahezu allen ausgelieferten WLAN Komponenten sind die Sicherheitsmechanismen standardmäßig deaktiviert. Es ist leicht verständlich, dass weder starke noch schwache Sicherheitsprotokolle schützen können, wenn sie nicht aktiviert werden. Kapitel 6.9 gibt dazu einige Hinweise. Insbesondere die verschiedenen im Einsatz befindlichen Sicherheitsprotokolle und Verfahren bergen einige Risiken, auf welche im Nachfolgenden eingegangen werden soll.

6.8.1 Mischbetrieb von WEP und TKIP

Die im Rahmen dieser Arbeit vorgestellten Verschlüsselungsprotokolle zum Schutz einer Funkverbindung, WEP und TKIP (WPA), können parallel verwendet werden. Insbesondere neuere Zugangspunkte sind in der Lage zu erkennen, welche Protokolle die mobilen Stationen verwenden. Sie können sich dann entsprechend anpassen, so dass ein Mischbetrieb möglich wird. Dieser Mischbetrieb birgt einige Gefahren [24]. Multicast und Broadcast Nachrichten werden nach wie vor mit WEP verschlüsselt, da mobile Stationen mit WEP die WPA Nachrichten nicht verstehen würden. Des weiteren muss davon ausgegangen werden, dass zu WPA inkompatible WLAN-Adapter auch keine Authentifizierung durch 802.1x unterstützen. Ein möglicher Angreifer kann auf diesem Weg die Authentifizierung umgehen. Mit dem nicht vorhandenen Schutz durch WEP kann also das gesamte Netz kompromittiert werden. Es ist darauf zu achten einen Mischbetrieb möglichst schnell

zu beenden und die Zugangspunkte so zu konfigurieren, dass sie dann auch keine WEP Verbindungen mehr akzeptieren.

6.8.2 Schlüsselmanagement

Größere Netzwerke sollten unbedingt über ein automatisches Schlüsselmanagement verfügen. Aus diesem Grund ist WEP für solche Netze nicht geeignet. Bei der Verwendung von WEP muss an jedem Zugangspunkt und jedem WLAN-Adapter der gemeinsame Schlüssel manuell eingegeben werden. Auch stärkere Protokolle, welche statische Hauptschlüssel ohne Authentifizierung verwenden, sind von diesem Problem betroffen. Wird beispielsweise ein tragbarer Computer, welcher für dieses WLAN konfiguriert ist, gestohlen, so ergeben sich fatale Folgen. Ein Angreifer kann den Hauptschlüssel des WLAN aus diesem Rechner auslesen und hat fortan freien Zugang zum Netzwerk (Im Fall von WEP ist der WEP-Schlüssel lediglich ein Eintrag in der Windows Registry). Der Administrator dieses WLAN müsste nun alle Schlüssel an den Zugangspunkte und allen WLAN Adaptern austauschen. Bei einem hinreichend großen WLAN ist das mit einem immensen Aufwand verbunden, was zur Folge haben kann, dass es nicht getan wird. Die Sicherheit in diesem Netz reduziert sich folglich auf Null.

Ein automatisiertes Schlüsselmanagement/Authentifizierungsverfahren löst dieses Problem und ist für große Netzwerke mit hohem Sicherheitsbedarf anzustreben.

6.8.3 Abschotten durch Firewall

Solange die Funkverbindung zwischen einem mobilen Rechner und dem Zugangspunkt als nicht sicher angesehen werden kann, ist es sinnvoll, die Funkverbindungen außerhalb einer Firewall zu halten. Abbildung 6.11 zeigt eine Anordnung ohne Firewall. In dieser

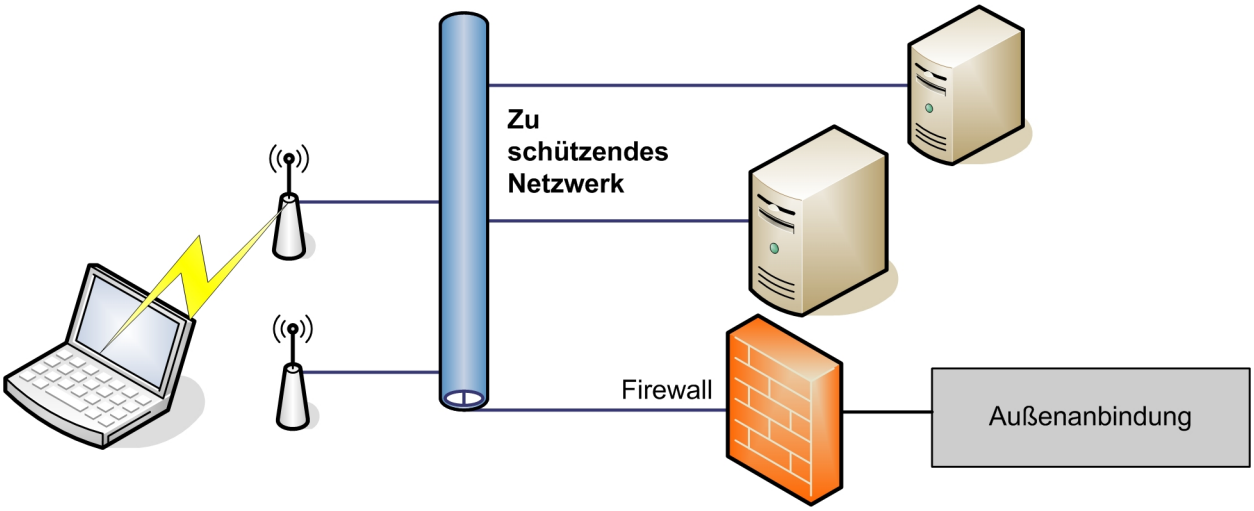

Abbildung 6.11: Netzwerkarchitektur ohne Firewall

Anordnung erlangt ein Angreifer Zugriff auf das gesamte Netz mit allen darin befindlichen Ressourcen, falls er es schafft den Schutz der Funkverbindung zu brechen. Abbildung 6.12 zeigt eine alternative Konfiguration [21]. Bei der Konfiguration nach Abbildung 6.12

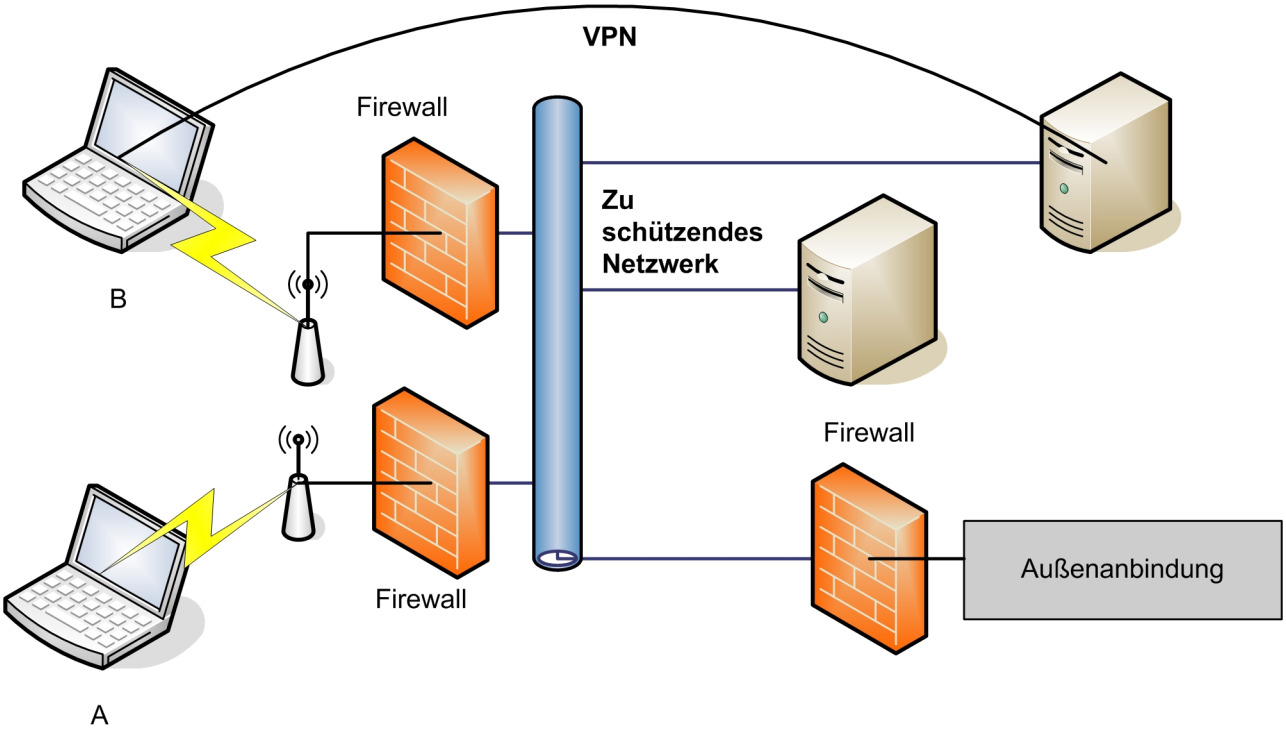

Abbildung 6.12: Netzwerkarchitektur mit Firewall

befinden sich die Zugangspunkte mit den Funkverbindungen außerhalb einer Firewall. Gelingt es einem Angreifer den Schutz der Funkverbindung zu brechen, so befindet er sich dennoch außerhalb des zu schützenden Netzwerkes. Die Firewall kann den Angreifer außerhalb des Netzes halten oder ihn durch Intrusion Detection[8] erkennen. Rechner A auf der Abbildung hat natürlich keinen Zugang mehr zu den Ressourcen des Netzes hinter der Firewall. Sollte ein Rechner Zugriff auf diese Ressourcen benötigen, so kann zum Beispiel ein VPN Tunnel durch die Firewall in das Netzwerk etabliert werden, wie die Abbildung 6.12 für Rechner B andeutet. Ist dieser Tunnel durch eine ausreichend starke Verschlüsselung geschützt, wird der entstehende Schaden bei einem Angriff auf die Funkverbindung zu Rechner B begrenzt. Je nach eingesetztem Sicherheitsprotokoll könnte der Angreifer z. B. Rechner A angreifen und vielleicht Zugriff auf das Internet erlangen. Zwischen dem Angreifer und dem zu schützenden Netzwerk stehen jedoch immer noch die starke VPN Verschlüsselung und die Firewall.

[8] Eindringlingserkennung

6.9 Maßnahmen zur Erhöhung der Sicherheit in Netzwerken

Wie die bisherigen Ausführungen zeigen, ist Sicherheit in kabellosen Netzwerken keine Selbstverständlichkeit. Sicherheit ist jedoch immer eine Frage des Aufwandes und der zu schützenden Daten und Ressourcen. Für Anwender mit hohem Sicherheitsbedarf kommen Funknetze entweder überhaupt nicht, oder nur mit umfassenden organisatorischen Maßnahmen und starken Verschlüsselungsalgorithmen in Frage.

Die starke Nachfrage und der wachsende Markt für Wireless Networks zeigen jedoch, dass ungeachtet der Sicherheitsprobleme, ein hohes Interesse an dieser Technologie besteht. In diesem Kapitel sollen einige Maßnahmen erläutert werden, welche das Aufspüren und Kompromittieren von Funknetzwerken erschweren und somit die eigene Sicherheit erhöhen [24].

- **Aktivieren der Schutzmaßnahmen**
  Aktivieren Sie die vorhandenen Schutzmaßnahmen ihrer Netzwerkkomponenten. Auch wenn WEP leicht zu brechen ist, erfordert es noch einen Aufwand und einige Computerkenntnisse.
- **SSID[9] abschalten oder umbenennen**
  Falls es technisch und organisatorisch möglich ist, sollte der SSID - Broadcast deaktiviert oder zeitlich begrenzt werden, so dass ein Firmennetzwerk z. B. nur in der Arbeitszeit aktiv ist. Sollte die SSID weiterhin ausgestrahlt werden, sollte sie umbenannt werden, so dass keine Rückschlüsse auf den Inhalt des Netzes gewonnen werden können
- **MAC Filterung einrichten**
  Die Zugangspunkte so zu konfigurieren, dass sie nur authorisierte MAC Adressen akzeptieren, erfordert einen gewissen Aufwand, steigert jedoch den Aufwand welcher erforderlich ist um in ein WLAN einzubrechen.
- **Statische Schlüssel periodisch wechseln**
  Werden statische Schlüssel in einem WLAN verwendet, so sind diese regelmäßig zu wechseln, um das WLAN sicher zu halten. Bei der Passwortwahl sind die entsprechenden Regeln — Länge und Zeichenvielfalt — zu beachten. Bei manchen Zugangspunkte sind Default-Passwörter für den Administrationszugriff eingestellt. Diese müssen unbedingt geändert werden. Zugangspunkte, welche nur durch den Kabelzugang konfiguriert werden können, bieten eine noch größere Sicherheit.
- **Technische Maßnahmen**
  Wenn möglich sollte die Leistung der Zugangspunkte so weit reduziert werden, dass nur der gewünschte Bereich Zugriff auf den Acess Point hat. Das Auffinden des Netzwerkes wird so erschwert.
- **DHCP abschalten**
  Wenn die automatische Vergabe von IPs nicht benötigt wird, sollte sie deaktiviert werden. Eine statische IP-Vergabe mit möglichst kleinem Adressraum verhindert eine automatische Zuweisung einer gültigen IP an einen Angreifer.

[9] SSID Service Set IDentifier: ermöglicht hier die Bennennung eines Funknetzwerkes

- **Firmware Upgrades durchführen**
  Wenn möglich sollten Updates von Treibern und Software auf den WLAN-Komponenten durchgeführt werden, da dies ebenfalls die Sicherheit erhöhen kann. Hierbei ist allerdings auf Kompatibilität der Komponenten untereinander zu achten.

6.10 Zusammenfassung

In den letzten Jahren fanden kabellose Netze eine starke Verbreitung. Allerdings wurde schnell klar, dass die ersten Sicherheitsprotokolle, insbesondere WEP, keinen nennenswerten Schutz bieten. Es wurden und werden Schritte unternommen, mehr Sicherheit für Funknetzwerke zu bieten. Mit der Einführung des Standards IEEE 802.11i wird die Sicherheit in Funknetzwerken einen großen Sprung nach vorne machen. Die AES-Verschlüsselung, auf welcher der Standard beruht, hat sich jahrelang bewährt und wird Angriffe auf kabellose Netzwerke massiv erschweren. Leider lässt der Standard im Moment noch auf sich warten. Aber auch die WPA Zwischenlösung mit TKIP und Authentifizierungsmaßnahmen steigert die Sicherheit im WLAN im Vergleich zum unsicheren WEP beträchtlich.

Ein Problem, welches Angreifern allerdings noch lange Zeit Tür und Tor öffnen wird, ist das mangelnde Sicherheitsbewusstsein vieler Betreiber von Funknetzwerken. Solange mögliche Sicherheitsmechanismen nicht aktiviert und konfiguriert werden, kann ein WLAN nicht sicher sein.

Totale Sicherheit wird es, auch im WLAN, vermutlich nie geben. Ein sicherheitsbewusster Nutzer hat mit WPA oder auch VPN und einer vernünftigen Konfiguration jedoch die Möglichkeit sein WLAN gut zu schützen. Mit dem Erscheinen von IEEE 802.11i und entsprechenden Geräten steigen diese Möglichkeiten weiter, sofern der Standard nicht gravierende Designfehler aufweist.

Literaturverzeichnis

[1] *Schily fordert mehr Sicherheit bei drahtlosen Netzen*
http://www.heise.de/newsticker/result.xhtml?url=
/newsticker/meldung/45828&words=Sicherheit

[2] Bundesamt für Sicherheit in der Informationstechnik
Studie "Durchführungskonzept für Penetrationstests", März 2004

[3] Kismet Homepage, März 2004
http://www.kismetwireless.net/index.shtml

[4] Computec Bugliste 2003/04/30
Kerio Personal Firewall durch Replay-Attacke angreifbar
http://www.computec.ch/buglist/2003-04/

[5] Hanns-Wilhelm Heibey, Ursula Meyer zu Natrup, Ralf Hauser und Carsten Schmidt, März 2004, *Datenschutz und Informationstechnische Sicherheit bei PC's*
http://www.datenschutz-berlin.de/infomat/pc/index.htm

[6] *Neues Loch in WLAN-Verschlüsselung*, März 2004-03-27
http://www.heise.de/newsticker/result.xhtml?url=
/newsticker/meldung/19928&words=WEP

[7] Jörg Rech, *Wireless LANs*, Heinz Heise Verlag, 2004

[8] Website der Firma RSA, März 2004, http://www.rsasecurity.com/

[9] *Funk-LAN-Verschlüsselung WEP passiv durchbrochen*, März 2004
http://www.heise.de/newsticker/result.xhtml?url=
/newsticker/meldung/20016&words=WEP

[10] Computer Science Division at the University of California Berkeley, März 2004
Security of the WEP algorithm
http://www.isaac.cs.berkeley.edu/isaac/wep-faq.html

[11] Scott Fluhrer, Itsik Mantin und Ado Shamir, März 2004
Weaknesses in the Key Scheduling Algorithm of RC4
http://www.drizzle.com/~aboba/IEEE/rc4_ksaproc.pdf

[12] Air Snort Homepage, März 2004, http://airsnort.shmoo.com/

[13] *Funknetze: Nachbesserung für mehr Sicherheit verspätet sich*
http://www.heise.de/newsticker/result.xhtml?url=
/newsticker/meldung/34811&words=802%2011i
WiFi NewsLetter, März 2004
http://wifi.weblogsinc.com/entry/7278529296341507/

[14] Wi-Fi-Homepage, Wi-Fi-Mitgliederliste, März 2004
http://www.wi-fi.org/OpenSection/index.asp
http://www.wi-fi.org/OpenSection/members.asp?TID=2

[15] *Verbesserung für WLAN- Sicherheit*, März 2004
http://www.heise.de/newsticker/result.xhtml?url=
/newsticker/meldung/31967&words=wpa%20WPA

[16] Wi-Fi Alliance, März 2004
Strong, standards-base, interoperable security for today's Wi-Fi network
http://www.wi-fi.com/OpenSection/pdf/Whitepaper_Wi-Fi_Security4-29-03.pdf

[17] Nancy Cam-Winget, Russ Housley, David Wagner und Jesse Walter, März 2004
Security Flaws in 802.11 Data Link Protocols
http://www.cs.berkeley.edu˜daw/papers/wireless-cacm.pdf

[18] Onn Haran, Olli Pekka Hiironen, März 2004
CTR Mode für Encryption
http://www.ieee802.org/1/linksec/meetings/PrevSep02/haran_p2mp_2_0702.pdf

[19] Vocal Technologies, März 2004, *CCMP AES Counter CBC-MAC Protocol Advanced Encryption Standard*, http://www.vocal.com/CCMP_AES.pdf

[20] IEEE Konsortium, März 2004
Port-Based Network Access Control 802.1x
http://standards.ieee.org/getieee802/download/802.1X-2001.pdf

[21] DELL, März 2004
White Paper Wireless Security in 802.11(Wi-Fi) Networks
http://www.dell.com/downloads/global/vectors/wireless_security.pdf

[22] Markus Nispel (Enterasys), März 2004
User Personalized Network EAP - Extensible Authentication Protocol
http://www.enterasys.com/de/products/whitepapers/EAP_Artikel.pdf

[23] CISCO, März 2004, *Public WLAN SIM Authentication und Authorization*
http://www.cisco.com/warp/public/cc/pd/witc/itp/prodlit/mapga_wp.pdf

[24] Bundesamt für Sicherheit in der Informationstechnik, März 2004
Sicherheit im Funk LAN
http://www.bsi.de/literat/doc/wlan/wlan.pdf

Kapitel 7

Intrusion Detection in mobilen Netzwerken

Lars Langer

Im letzten Jahrzehnt gewann die Intrusion Detection fortwährend an Bedeutung. Parallel dazu keimte die mobile Umgebung für Netzwerke auf. Deshalb entstanden für die Intrusion Detection immer neuere Herausforderungen und Probleme, die gelöst werden mussten. Mit einigen Lösungsansätzen stehen den Netzwerkbetreibern auch hierfür bereits einige Möglichkeiten offen, mit Intrusion Detection ihre mobilen Netzwerke und Endgeräte sicherer zu gestalten. Mit der zunehmenden Effektivität von Intrusion Detection Systemen, nimmt auch die Zahl der Interessenten zu, die mit dieser zweiten Verteidigungslinie beabsichtigen, sich selbst oder ihren Kunden einen höheren Grad an Sicherheit zu gewährleisten. Trotz dessen wird Intrusion Detection immer nur eine Ergänzung zu auch anderweitig geschützten Netzwerken bleiben, und niemals die Firewall ersetzen können.

Inhalt dieser Arbeit soll sein, auch dem weniger sachkundigen Leser nahe zu bringen, was Intrusion Detection leisten kann, indem zunächst die dazu angewandten Methoden skizziert werden gefolgt von den Möglichkeiten ihren Einsatz zu variieren. Wichtig ist ebenso, darzulegen welche Ansätze dem Benutzer bereits geboten werden Intrusion Detection in drahtloser Umgebung einzusetzen und damit verbundene Herausforderungen und Probleme zu bewältigen.

Inhaltsverzeichnis

7.1 Einleitung 149
7.2 Intrusion Detection und angewandte Methoden 149
7.2.1 Allgemeines 149
7.2.2 Anomaly Detection 150
7.2.3 Signature Detection 152
7.2.4 Compound Detection 153
7.2.5 Arten der Angriffe 154
7.3 Möglichkeiten des Einsatzes von Intrusion Detection 154
7.3.1 Netzwerkbasiert und Rechnerbasiert 155
7.3.2 Online und Offline Modus 156
7.3.3 Intrusion Response 156
7.3.4 Umgang mit Prüfdaten 157
7.3.5 Weitere Aspekte 158
7.3.6 Trends und Ziele 158
7.4 Probleme und Lösungen in drahtloser Umgebung 159
7.4.1 Grundsätzliche Probleme 159
7.4.2 Herausforderungen 160
7.4.3 Vorgehensweise in Infrastruktur-Netzwerken 160
7.4.4 Vorgehensweise in Ad Hoc-Netzwerken 161
7.5 Fazit 162

7.1 Einleitung

Mobile Netzwerke breiten sich immer mehr aus. Damit sie auch in sicherheitsempfindlichen Bereichen eingesetzt werden können, müssen sie ähnlich hohen Sicherheitsanforderungen genügen wie drahtgebundene Netze. Ein wichtiges Werkzeug um die Integrität eines Netzwerkes sicher zu stellen sind Intrusion Detection Systeme. Sie überwachen ständig den Datenverkehr und versuchen Anomalien und Angriffe aufzuspüren und daraufhin angemessene Reaktionen einzuleiten. Diese Arbeit stellt das Konzept von Intrusion Detection Systemen vor und zeigt mögliche Klassifizierungen auf. Anhand von auf dem Markt befindlichen Lösungen stellt die Seminararbeit weiterhin dar, in wie weit diese auch für drahtlose Netzwerke einsetzbar sind.

7.2 Intrusion Detection und angewandte Methoden

In diesem Kapitel wird nach einem kurzen allgemeinen Überblick auf die angewandten Methoden der Intrusion Detection eingegangen und im folgenden seine Anwendung auf bestimmte Arten von Angriffen vorgestellt.

7.2.1 Allgemeines

In den letzten Jahren werden Informationssysteme zunehmend vernetzt. Da sich die Bedrohungslage im Internet, aber auch innerhalb eines Firmennetzwerkes (Intranet, LAN usw.) verschlechtert, erfordert die Systemsicherheit ein höheres Maß an Berücksichtigung. Kein noch so fortschrittliches System kann einen lückenlosen Schutz garantieren. Deshalb finden neben bereits obligatorisch eingesetzten Firewalls auch Intrusion Detection Systeme (IDS) immer weitere Verbreitung. Doch was ist Intrusion Detection (ID) überhaupt? Hierbei handelt es sich um die Möglichkeit einer Zweite Linie der Verteidigung hinter derjenigen der Firewalls und Zugangskontrollen etc. aufzubauen. Man möchte etwas unternehmen können, wenn die bisherigen Sicherheitsvorkehrungen umgangen worden sind. Intrusion Detection bietet die Möglichkeit unautorisiertes Modifizieren von Daten eines beobachteten Systems oder des Datenaustauschs zwischen mehreren Systemen aufzuspüren und angemessene Maßnahmen zu ergreifen. Problematisch ist außerdem, dass die Bedrohung das System nicht nur von Außen durch Hacker, Script Kiddies, Terroristen, Spione oder bösartigem Code von Malware gefährdet ist, sondern auch von Innen. Gemeint sind damit Mitarbeiter, die bösartig oder durch Erpressung versuchen dem System Schaden zuzufügen, die vielleicht eine mangelnde Ausbildung erhalten haben oder durch die sich in der Konfiguration des Intrusion Detection Systems Fehler eingeschlichen haben. Intrusion Detection Systeme sind ein relativ junges Forschungsgebiet, zu dem es kaum herstellerunabhängige Daten gibt und es auch trotz der enormen Fortschritte im letzten Jahrzehnt noch immer nicht den angemessenen Einzug in die Lehrbücher geschafft

hat. Anfangs war Intrusion Detection auch nur ein Versuch die Auditdaten eines Mainframes auf ein kompaktes aussagekräftiges Niveau zu reduzieren[2]. Auditdaten ist der Begriff, mit dem man bei Intrusion Detection Systemen generierte Prüfdaten beschreibt. In der ID versucht man insbesondere mit 2 Methoden nach einem eventuellen Eingriff zu suchen. Im Folgenden wird auf diese Methoden spezieller eingegangen.

7.2.2 Anomaly Detection

Bei dieser Methode geht man von der Annahme aus, dass eine Bedrohung, sei es ein Angreifer, Malware oder andere unautorisierte Modifikation der Systemdaten, durch seine Aktionen das Verhalten des Systems so beeinflusst, dass es von seinem normalen Verhalten abweicht. Aufgabe dieser Methode ist nun diese Abweichungen aufzuspüren und darauf entsprechend zu reagieren. Die Vorgehensweise ist, dass vom System Messdaten gesammelt werden und zu statistischen Profilen (behaviour profiles) zusammengefasst werden. Hierbei besteht auch die Möglichkeit der Eingrenzung der Detection auf ein Teilsystem oder einen User. Im Folgenden wird darauf eingegangen, dass diese Vorgehensweisen sowohl programmiert als auch vom ID System selbst erlernt sein können[1].

Selbstlernende Systeme

Diese Systeme werden in drei Kategorien unterschieden. Das Unterscheidungskriterium sind die Mittel der Beschreibung des gutartigen Verhaltens.

Rule Modelling:

Hierbei wird zunächst der auftretende Verkehr der Daten studiert. Es folgt der Versuch die normalen Operationen durch das Aufstellen von Regeln zu beschreiben. Bei einem Regelverstoß wird Alarm ausgelöst. Auf mögliche weitere Reaktionen des Systems wird im Kapitel 3.3 genauer eingegangen.

Descriptive Statistics:

Nach einer Sammlung von einfachen Grenzwerten innerhalb der systemtypischen Operationen wir daraus ein Mittelwert gebildet. Daraufhin folgt die Berechnung der Standardabweichungen. Schließlich kontrolliert man durch einen daraus gebildeten Abstandsvektor (distance vector) die Abweichungen, die im System vorkommen. Bei einer Distanz, die nicht im Toleranzbereich liegt, wird ein Alarm ausgelöst.

Künstliches neurales Netz:

Das ANN - Netz (artificial neural network) wird zunächst mit dem normalen Verkehr gefüttert. Daraus versucht es Strukturen und Muster (pattern) zu deuten und zu erkennen. Beim Einsatz des ANN wird das aktuelle Muster mit den studierten verglichen. Sind die Strukturen verschieden kommt es zum Alarm.

Programmierte Systeme

Programmierte Systeme werden in zwei Kategorien unterschieden. Der gravierende Unterschied liegt bei der Vorgehensweise zur Umschreibung von akzeptablem Verhalten.

Statistics:

Man bildet ein statistisches Profil des normalen Verhaltens durch die Speicherung und Sammlung einfacher Statistiken. Möglichkeiten hierfür wären die Anzahl von fehlgeschlagenen Logins, die Anzahl der Netzwerkverbindungen oder auch die Anzahl von Kommandos mit negativer Antwort (error return). Es bietet sich an, Alarm zu geben, wenn die Anzahl einen festgelegten Wert überschreitet.

Default Deny:

Die Idee davon ist, die Umstände, unter denen das System als gutartig eingestuft wird, also alle sicherheitskonformen Zustände, explizit anzugeben. Damit entsteht eine klare Übereinstimmung mit einer Sicherheitspolitik mit vorgegebener Ablehnung jeglichen anderen Verhaltens.

Diese Vorgehensweisen sind bzw. enthalten strikte Anweisungen für das System, was ein normales Verhalten sein darf. Jedwede Abweichung von diesem erlaubten Verlauf ist inakzeptabel und führt zu einem Alarm. Problematisch hierbei ist es, dass eine Abweichung vom System kein bösartiges Verhalten sein muss, sondern lediglich Aktionen von neuem normalem Verhalten sein könnten. Deshalb kommt es in Anomaly Detection Systemen zu einer nicht unerheblichen Anzahl von Fehlalarmen.

Überblick

Dieser Überblick stellt kurz das Pro und Contra gegenüber.

Vorteile:

Anomaly Detection Systeme sind unabhängig von einer Signaturdatenbank. Deshalb ist es in der Lage alle möglichen Angriffe zu erkennen, ganz gleich, ob der Angriff bereits bekannt ist, oder auf einer neuen Taktik beruht.

Nachteile:

Der rechentechnische Aufwand ist höher als bei der folgenden Methode (Signature Detection). Darunter leidet die Performance des Systems. Die Fehlalarmrate dieses Systems ist sehr hoch, und neue Aktivitäten können nicht von Angriffen und anderen Anomalien Unterschieden werden. Aufgrund dieser hohen Anfälligkeit für (Fehl)Alarme sind Täuschungen und Denial of Service Angriffe mit geringem Aufwand möglich. Für den Administrator ist diese Tatsache sehr arbeitsaufwendig und wenig tragbar.

7.2.3 Signature Detection

Diese Methode geht davon aus, dass ein bestimmter Angriff nach einem ganz bestimmten Muster folgt. Verschiedenartige Angriffe haben dementsprechend verschiedene Muster. Beispielsweise verursacht eine synchronization-flood Attacke viele unbeantwortete TCP Synchronisationspakete. Solche Strukturen, genannt Angriffssignatur (attack signature), versucht das Signature Detection System in den Auditdaten zu finden. Vorraussetzung dafür ist jedoch, dass der Angriff bereits zuvor bekannt geworden ist, und seine Signatur in der Datenbank des gegenwärtigen ID Systems enthalten ist. Im Bereich der Signature Detection Methoden gibt es derzeit noch einen großen Nachholbedarf bei den selbstlernenden Systemen. Es werden derzeit nur programmierte Vorgehensweisen angewandt, die im Folgenden erklärt werden[1].

Programmierte Systeme

Programmierte Systeme in der Signature Detection können in vier Kategorien unterschieden werden. Im Gegensatz zu Anomalie Detection geht es hierbei um die Unterscheidung der Mittel zur Beschreibung des bösartigen Verhaltens.

State Modelling:

Bei dieser Vorgehensweise existieren mehrere Zustände, die einen Alarm auslösen können. Wichtig dabei ist, dass jeder dieser Zustände im beobachteten Bereich existent gewesen sein muss, damit man mit Sicherheit feststellen kann, dass ein Eingriff stattgefunden hat. Die Anordnung der Zustände kann im günstigsten Fall als einfache Kette ablaufen, es sind jedoch auch verzweigtere Gebilde möglich, beispielsweise Baumstrukturen oder Petrinetze.

Expert System:

Die Entscheidung eines Experten Systems basiert auf einer Menge von komplexen gegebenen Regeln. Diese Vorgehensweise ist am angemessensten, wenn im System ständig neue Fakten und zu prüfende Ereignisse stattfinden. Dadurch, dass es das Experten System auch ermöglicht selbstmodifizierenden bösartigen Code eines Angriffs aufzuspüren, ist es eine der mächtigsten Möglichkeiten der Intrusion Detection. Weiterhin bietet es durch seine Regelbasiertheit einen hohen Grad an Flexibilität. Doch diese Vorteile fordern auch einen Preis. Experten Systeme verzeichnen hohe Verluste in der Performance des kontrollierten Systems, nicht zuletzt aufgrund der Komplexität der Regeln.

Simple rule based:

Diese Systeme ähneln den mächtigeren Experten Systemen. Da hier mehr Wert auf die Performance des Systems gelegt wurde, verzichtet man darauf, solch komplexe Regeln zu verwenden.

String Matching:

Hierbei handelt es sich um ein Verfahren, das die Bitfolgen bzw. den Text der zwischen mehreren Systemen übermittelt wird zu durchsuchen um eventuelle kleinere Bitfolgen und Strings darin zu finden, die einem Angriff aus der Datenbank zugeordnet werden können.

Bei Signature Detection Systemen sind Abweichungen, die keinem bekannten Angriff entsprechen, akzeptabel. Es soll darauf hingewiesen sein, dass man in anderer Literatur möglicherweise den Begriff Misuse Detection finden kann, der dem Begriff der Signature Detection entspricht.

Überblick

Hier werden Vor- und Nachteile zusammenfassend dargestellt.

Vorteile:

Ein Signature Detection System ist im Allgemeinen recht schnell und findet zuverlässig bekannte Angriffssignaturen. Dadurch gilt es weithin als das stärkere der beiden Methoden. Besonders im Bereich der Online-Überwachung des Verkehrs in Netzwerken wird vorwiegend auf Signature Detection zurückgegriffen. Außerdem kommt es bei fehlerfreier Analyse der Auditdaten zu keinem Fehlalarm, da nur bekannte Angriffssignaturen Alarm auslösen können.

Nachteile:

Da die aufzuspürenden Angriffe komplett bekannt und bereits analysiert sein müssen, damit die attack signature in der Datenbank zur Verfügung steht, ist diese Methode abhängig von dieser Datenbank und von ständigen Updates. Unbekannte und neue Angriffe können bei Signature Detection gefährlich werden, da sie nicht erkannt werden können.

7.2.4 Compound Detection

Hierbei handelt es sich lediglich um ein System, dass von der Methode der Signature Detection abgeleitet wurde. Der Begriff 'signature inspired' ist daher auch häufig diesbezüglich in anderer Literatur zu finden[1]. Tatsächlich jedoch zieht dieses Verfahren beide Methoden zu Rate, sowohl Anomaly Detection als auch Signature Detection. Eine mögliche Anwendung wäre wie folgt[4]. Dieses ID System sucht im beobachteten Bereich nach Anomalien. Sobald es solche Abweichungen vom normalen Verhalten gibt, werden die betroffenen Daten nach attack signatures durchsucht. Somit können auch unbekannte Angriffe einen Alarm auslösen, und es besteht die Möglichkeit verschiedene Alarmstufen zu integrieren. Dieses Verfahren bietet demnach eine deutlich bessere Chance wirklich interessante Ereignisse zu entdecken, da sowohl ein Vergleich zwischen aktuellem Verhalten und normalem Verhalten, als auch zwischen aktuellem Verhalten und bösartigem Verhalten gemacht wird. Andererseits besteht auch die Möglichkeit durch das Anwenden beider Methoden eine bessere Abstimmung zwischen normalem und bösartigem Verhalten zu erzielen, und in Strukturen oder Regeln festzuhalten. Üblicherweise sind solche Verfahren selbstlernend[1].

7.2.5 Arten der Angriffe

Um kategorisieren zu können bei welchen Angriffen welche Methode geeignet scheint wird nun vorgestellt, in welche drei Arten Angriffe unterschieden werden können[1]. Im Folgenden sind diese Arten danach geordnet, welche Schwierigkeit sie beim Detektieren bereiten.

Well Known Intrusions:

Angriffe dieser Art sind sehr leicht zu finden. Sie sind bereits gut bekannt und demnach sind deren Angriffsstrukturen in jeder aktuellen Datenbank eingetragen. Der Grund dafür liegt einmal in der Einfachheit der Angriffssignatur, die zudem von Angriff zu Angriff auch kaum oder keine Variationen benutzen. Zum anderen liegt es daran, dass Angriffe dieser Art meist ein ganz bestimmtes Sicherheitsloch benutzen, wodurch die Lokalisierbarkeit erheblich vereinfacht wird, ebenso die Abwehr durch eventuelle Updates, die dieses Loch schließen. Aufgrund der gegebenen attack signature wird hierfür Signature Detection bevorzugt.

Generalizable Intrusions:

Diese Eingriffe ins System kann man mit etwas mehr Aufwand auch noch finden. Sie haben häufig komplexere Variationen, beispielsweise in Form von variierenden Petrinetzen, und benutzen nicht nur ein Sicherheitsloch, sondern greifen durch mehrere auf einmal an. Das macht sie etwas schwerer zu entdecken, aber mit Compound Intrusion Detection und kombinierter Intrusion Detection bestehen gute Chancen, dass man sie aufspüren kann um darauf zu reagieren.

Unknown Intrusions:

Es kann vorkommen, dass diese Angriffe noch nicht ausreichend studiert worden sind, oder komplett unbekannt sind. Solch ein Angriff nutzt in der Regel ein Sicherheitsloch, das zuvor noch nicht gefunden oder nicht weiter beachtet wurde. Dadurch sind sie sehr schwer zu finden. Lediglich Anomalien können auf ihre Existenz hinweisen, weshalb man auch nur Anomaly Detection gegen diese Angriffe einsetzen kann.

7.3 Möglichkeiten des Einsatzes von Intrusion Detection

Dieses Kapitel beschäftigt sich damit, auf welche Art ein Intrusion Detection System eingesetzt werden kann.

7.3.1 Netzwerkbasiert und Rechnerbasiert

Ein ID System kann zur Überwachung von Netzwerken oder Rechnern verwendet werden. Im Folgenden wird auf diesen Unterschied genauer eingegangen.

Netzwerkbasiert:

Ein Intrusion Detection System wird in diesem Fall dazu verwendet, ein Netzwerk zu überwachen[1]. Diese Anwendung ist für einen Angreifer unsichtbar, da das System lediglich die anfallenden Daten abhören muss. Man spricht deshalb auch davon, dass netzwerkbasierte ID im 'Stealth-Modus' operieren. Trotz ihrer Aufgabe den gesamten Netzwerkverkehr zu überwachen, besteht auch hier die Möglichkeit, die Verfahren einer zentralen Kontrolleinheit unterstellen zu können. Üblicherweise versucht man die Auditdaten, falls möglich, an einem Engpass wie einem Router, Gateway etc. abzugreifen und der Prüfung zu unterziehen. Da gerade in einem Netzwerk eine Vielzahl verschiedener Vorgänge und Zustände auftreten, nicht zuletzt weil mehrere Systeme das Netzwerk benutzen, bietet sich Anomaly Detection weniger an und man setzt statt dessen fast immer auf Signature Detection. Das zieht natürlich nach sich, dass bei der Erkennung neuer Angriffe unweigerlich Probleme auftreten. Das erfordert das ständige Aktualisieren der Signature-Datenbanken über das Netz. Auch diese können während eines Angriffs in das Netzwerk in Mitleidenschaft gezogen werden. Konflikte gibt es derzeit auch mit kryptographisch geschützten Verbindungen, da die aktuellen Daten nur in Klartextform mit den vorliegenden Daten der Signaturdatenbank verglichen werden können. IDS schauen hierbei nicht nur in die Protokollköpfe, sondern auch in die Nutzdaten. Da dies bei verschlüsselten Verbindungen nicht möglich ist, kann das Intrusion Detection System die Daten nicht überwachen. Es sei darauf hingewiesen, dass kryptographisch geschützte Verbindungen auch weniger anfällig für Angriffe sind, da sie durch die Verschlüsselung selbst für einen höheren Grad an Sicherheit sorgen. Die verschlüsselten Daten können normalerweise nur vom Sender und Empfänger genutzt werden. Dadurch wird dieses Problem eher nebensächlich. Nur bei internen Bedrohungen kann diese Situation zu Gefahr werden. Man kann auch nicht davon ausgehen, hier generelle Lösungen zu finden, da die Verschlüsselung ja darauf ausgelegt ist, dass sie niemand fremdes entziffern kann, auch nicht das ID System.

Rechnerbasiert:

Im Gegensatz zum netzwerkbasierten Intrusion Detection System was den Verkehr verschiedener Rechner kontrolliert, beschäftigt sich hier das ID System mit dem Prüfen der Auditdaten eines einzelnen Systems, Teilsystems oder Users[1]. Die Durchsuchung läuft im Wesentlichen in den sicherheitsrelevanten Protokolldateien des Rechners ab. Beispielsweise versucht man Eindringlinge in Kernel Logfiles, Router Logs, Firewall Logs oder Logfiles von Anwendungsprogrammen aufzudecken. Die Integrität dieser und ähnlicher systemrelevanter Daten erhalten hier die höchste Priorität.

7.3.2 Online und Offline Modus

Hier wird unterschieden, ob das IDS permanent oder sporadisch auf die zu prüfenden Daten angewendet wird.

Online:

Ein ID System läuft online, wenn eine permanente Überwachung des aktuellen Verkehrs vollzogen wird[1]. Hierbei unterscheidet man zwischen zwei Arten der Online-Überwachung. Die erste ist die Kontinuierliche. Das heißt, dass die produzierten Auditdaten unmittelbar der Prüfung auf Angriffe oder Anomalien unterzogen werden, ohne vorher zwischengespeichert zu werden. Die zweite Art läuft gestapelt ab. Hierbei werden innerhalb kurzer Perioden die anfallenden Auditdaten zwischengespeichert und nach Ablauf dieser Perioden gesammelt der Prüfung unterzogen. Die Periodendauer ist dabei maximal einige Sekunden lang.

Offline:

Hier geht es darum die Logfiles und Protokolle des Systems oder eines Users jeweils nach einer längeren Arbeitsphase, beispielsweise einmal pro Tag, auf Einträge zu untersuchen, die einem Eingriff in das System entsprechen könnten[1]. Nachteil davon ist, dass man nicht die Möglichkeit hat zu reagieren und Schadensbegrenzung zu betreiben. Es ermöglicht aber gleichzeitig, dass die Logfiles gegebenenfalls mehrfach durchsucht werden können, um komplexere und verteiltere Angriffe lokalisieren zu können.

7.3.3 Intrusion Response

Es handelt sich bei Intrusion Response um die Maßnahmen des Systems, die unternommen werden können, wenn ein Angriff oder eine Anomalie erkannt worden ist. Problematisch ist dabei, dass Anomalien möglicherweise auch neue normale Aktivitäten sein können, da es im Bereich von Anomaly Detection zu häufigen Fehlalarmen kommen kann. Diesbezüglich sind die Studien noch sehr rar und erlauben derzeit noch keine akzeptable Fehlalarmbehandlung bzw. -vermeidung. Im Folgenden wird auf die Maßnahmen eingegangen die das Intrusion Detection System bei einem Eindringen in das System ergreifen könnte[1].

Passiv:

Ein passives ID System wird bei Alarm den Administrator verständigen, und ihm bei Bedarf die Informationen zukommen lassen, die das System über den Angriff in Erfahrung bringen konnte. Maßnahmen um gegen den Angriff vorzugehen werden nicht getroffen.

Aktiv:

Maßnahmen bezüglich des angegriffenen Systems:

In diesem Fall wird versucht, über das eigene System die Kontrolle zurück zu gewinnen. Im Allgemeinen ist man bestrebt die Auswirkungen des Angriffs zu vermeiden oder zumindest so gering wie möglich zu halten. Hierzu übliche Aktionen sind die Unterbrechung der Verbindung, beispielsweise durch das senden von TCP Reset Paketen, die Sperrung von kompromittierten Benutzerkonten, die möglicherweise infiltriert wurden um sich von dort aus Administratorrechte zu verschaffen, das Beenden schädlicher Prozesse, die eventuell Würmer oder auch Trojaner zugehörig sind, und das Aktivieren einer höheren Sicherheitsstufe, was neben anderen Maßnahmen durch eine Rekonfiguration der Firewalls geschehen kann.

Maßnahmen bezüglich des angreifenden Systems:

Hierbei wird versucht, in die Operationsplattform des Angreifers einzudringen und diese unschädlich zu machen und möglicherweise entwendete Daten zu entfernen. Diese Vorgehensweise ist rechtlich problematisch und wird deshalb nicht von Unternehmen und privaten Netzwerkbetreibern angewandt. Sie findet daher nur Verwendung beim Militär, beim Bundesnachrichtendienst, bei der Polizei oder anderen Einrichtungen des Bundes.

7.3.4 Umgang mit Prüfdaten

Hauptunterscheidungsmerkmal sind beim Umgang mit den Prüfdaten die Aspekte der Zentralisierung und Dezentralisierung.

Möglichkeiten der Datensammlung:

Auditdaten können an zwei verschiedenen Stellen gesammelten werden[1]. Sie können von einem einzelnen System, Teilsystem oder User gesammelt und kontrolliert werden, oder an einem Router, Gateway oder einer anderer zentralen Einheit abgegriffen werden, womit mehrere Systeme miteinander verbunden werden und worüber die Kommunikation gebündelt wird, so dass es nicht umgangen werden kann.

Möglichkeiten der Datenprüfung:

Ähnlich der Datensammlung kann auch die Datenprüfung in zwei Arten unterteilt werden. Die Auditdaten können einmal auf dem System geprüft werden, auf dem sie gesammelt werden, oder man verteilt diese Prüfdaten auf mehrere Intrusion Detection Systeme, die miteinander kooperieren[1]. So kann man weiterhin eine gute Performance des eigentlichen Systems gewährleisten, da die Prüfung selbst dieses System nicht belastet. Möglich ist auch, dass man die Ergebnisse bei dezentralisierter Datenprüfung miteinander abgleicht, also beispielsweise die anderen Systeme danach fragt, ob sie dieselbe Anomalie entdeckt haben, wie das eigene System.

7.3.5 Weitere Aspekte

Hierbei handelt es sich um bisher noch nicht beachtete Eigenschaften. Es wird auf die Interoperabilität eingegangen und eine Bewertung des Sicherheitsstandes versucht.

Interoperabilität:

Einige Intrusion Detection Systeme sind in der Lage mit anderen Intrusion Detection Systemen zu kooperieren, auch wenn möglicherweise zwei verschiedene Methoden der ID angewandt werden[1]. Eine günstige Variante wäre beispielsweise, ein Anomaly Detection System laufen zu lassen, was bei Aufdeckung einer Anomalie die Auditdaten an ein Signature Detection System weiterleitet. So können sowohl bereits bekannte Angriffe also auch unbekannte Angriffe detektiert werden.

Sicherheit:

Eine akzeptable Leistung bezüglich der Sicherheit gewährleisten ID Systeme seit höchstens 5 Jahren, davor waren die Systeme noch in Versuchsstadien oder basierten auf unzureichenden Analysen und Studien.

7.3.6 Trends und Ziele

Es soll kurz dargelegt werden, inwiefern welche Aspekte der in diesem Kapitel dargestellten Einsatzmöglichkeiten in der näheren Zukunft eine größere Rolle spielen werden. Man wird mehr und mehr von einem passiven Intrusion Response System auf aktive Systeme übergehen. Der Vollzug dieses Prozesses ist zum gegenwärtigen Zeitpunkt bereits gut fortgeschritten[1]. Wichtig ist dabei, dass eine sofortige Reaktion erfolgen soll, um den Angriff möglichst frühzeitig bekämpfen und Schäden vielleicht komplett vermeiden zu können. Welcher Prozess hingegen erst durch die drahtlosen Netzwerke verstärkt angeregt wurde, ist der Trend zu dezentralisierten System bezüglich der Datenprüfung und -sammlung[5]. Hier befinden wir uns derzeit im Anfangsstadium. Ein eindeutiger und auch zu erwartender Trend ist im Bereich kooperativer ID Systeme zu erkennen[1]. Fast jedes moderne Intrusion Detection System bietet die Möglichkeit zur Interoperabilität. Auch das allgemeine Interesse im Bereich der zweiten Verteidigungslinie gegen Angriffe steigt und wird vermehrt eingesetzt[5]. Schließlich ist es unumgängliches Ziel, Intrusion Detection Systeme ständig zu verbessern. Insbesondere zielt man darauf ab, dass einerseits die Angriffsmuster und Modelle des normalen Verhaltens verfeinert und verbessert werden, und andererseits, die auftretenden Fehlalarme zu minimieren und an den betreffenden Analysen zu forschen.

7.4 Probleme und Lösungen in drahtloser Umgebung

Nach einer Erläuterung der veränderten Situation durch drahtlose Netzwerke geht es in diesem Kapitel um spezielle neue Herausforderungen, die mit der mobilen Umgebung entstehen. Der wichtigste Teil hierbei sind die möglichen Lösungsansätze für dabei entstandene Probleme sowohl im Infrastruktur-Netzwerk als auch in Ad Hoc-Netzwerk.

7.4.1 Grundsätzliche Probleme

Das offene Medium:

Da elektromagnetische Wellen prinzipbedingt sich in jede Richtung ausbreiten, entsteht eine deutlich höhere Empfänglichkeit für Angriffe. Es wird möglich ohne direkten Kontakt, also ohne physikalischen Zugriff auf Netzwerkkabel, optische Fasern usw., sich in das Netzwerk einzuklinken. Auf allen Schichten des ISO/OSI-Modells (Layer), auf jedes Mobile Node (MN) und von jedem MN kann dadurch ein Angriff erfolgen[5]. Ein Angreifer kann mit geringerem Aufwand in eine internere Ebene vordringen und somit erheblichen Schaden anrichten. Er kann den Betroffenen auch passiv schädigen, indem er unbemerkt das Netz abhört. Die meisten Intrusion Detection Systeme brauchen statische Umgebungen. Daher kann es bei drahtloser und sich ständig ändernder Umgebung notwendig werden, ein neues Konzept für die Intrusion Detection in dynamischer Umgebung zu erarbeiten und anzuwenden.

Anfälligkeit der Software:

Um solche Angriffe durchzuführen werden Löcher und ungewollt kooperatives Verhalten von sicherheitsrelevanten Dateien und Protokollen ausgenutzt. Besonders MAC Protokolle sind sehr anfällig für Denial of Service Attacken[6], da nur simuliert werden muss, dass bestimmte Verbindungen bereits in Benutzung sind, und somit nicht für den angeforderten Service zur Verfügung stehen.

Weitere Gefahren:

Ferner sind auch statische Netzwerke durch unzureichend geschützte MNs gefährdet. Möglicherweise könnte ein Wurm in ein MN gelangen was sich ungeschützt im Internet aufhält. Wenn dieses MN sich nun in ein nach außen gut geschütztes statisches Netzwerk einklinkt, beispielsweise ein Mitarbeiter verbindet sich in seiner Firma mit dem LAN, so kann es zur Infizierung des statischen Netzwerkes von innen heraus kommen.

7.4.2 Herausforderungen

Dezentralisierung und Lokalisierung:

Die Möglichkeit des Angriffs ohne festes Medium erschwert beträchtlich die Lokalisierung des Angreifers. Andererseits kann die schlechte Lokalisierung auch zum Vorteil genutzt werden, wenn ein Angreifer ein ganz bestimmtes Ziel hat, da er dieses vorher auch erst einmal lokalisieren muss. Weiterhin gibt es in mobiler Umgebung weniger Verkehrskonzentrationspunkte, so dass die zu überprüfenden Daten von überall her kommen, und schließlich auch nur lokal vorliegen. Es stellt sich die Frage, ob man möglicherweise auch in dezentralisierter und globaler Form Intrusion Detection betreiben kann. Dies erfordert jedoch die Mitarbeit der anderen MNs im Netzwerk.

Daten für ID Systeme:

In drahtloser Umgebung werden neue Anwendungen und Verkehrsmuster nötig. Doch laufende Operationen können durch die zerbrechlichere Kommunikation leicht unterbrochen werden. Möglicherweise kann die Funkreichweite kurzzeitig nicht ausreichen oder das Signal trifft auf ein Hindernis. Das kann zwei Konflikte zur Folge haben. Wenn die Daten mit größerer Verzögerung eintreffen, so ist es für die üblicherweise echtzeit-basierten Intrusion Detection Systeme schwer bzw. nicht möglich, auch in Echtzeit die Daten zu prüfen[5]. Noch problematischer wird es, wenn die Daten unvollständig vorliegen, das heißt, wenn Daten auf dem Weg unwiederbringlich verloren gehen. Die Intrusion Detection Methoden können nur auf vollständige Daten angewandt werden, denn möglicherweise fehlte gerade ein solcher Teil, mit dem man eine Signatur oder Anomalie hätte erkennen können[5]. Da dann jedoch die Daten des möglichen Angriffs nicht vollständig sind, besteht weniger Gefahr für das überwachte System, man kann jedoch auch nicht in Erfahrung bringen, ob man von jemandem als Ziel auserkoren wurde. Es stellt sich die Frage, welche Algorithmen man bei unvollständigen Daten anwenden könnte.

7.4.3 Vorgehensweise in Infrastruktur-Netzwerken

In Infrastruktur-Netzwerken gibt es einige Lösungsansätze die im Folgenden genauer skizziert werden.

Probleme und Lösungsansätze bezüglich der Schichten des ISO/OSI-Modells:

In drahtloser Umgebung ist es für den Angreifer möglich auf allen Schichten zu operieren. Die MAC-Ebene ist hierbei eine besonders anfällige Schicht. Um das Problem zu bekämpfen, versucht man sie durch eine auf Layer 3 basierte Firewall zu isolieren[5]. Wenn das Intrusion Detection System eine Anomalie auf einem der unteren Layer entdeckt, so weist das häufig darauf hin, dass auf einem höheren Layer ein Eingriff stattgefunden hat.

Lösungsansatz für Angriffe auf Infrastruktur-Netzwerke:

Die Idee ist, dass jedes neue Mobile Node, welches sich ins Netzwerk einklinkt, sofort an einem gemeinsamen, globalen Intrusion Detection Prozess teilnimmt[5]. Nachdem es in den Prozess integriert wurde, fängt das lokale ID System an, Wissen über die Umgebung zu sammeln, um mit anderen Intrusion Detection Sensoren kooperieren zu können. Entscheidend ist dabei die umgehende Aktualisierung der Signaturdatenbank. Bei jedem Neuaufbau einer Verbindung (reconnect) ist dieser Vorgang zu wiederholen. Wichtig ist bei dem Integrationsprozess, dass er in der Lage sein sollte automatisch ablaufen zu können, da bei manuellem Einstellen damit zu rechnen ist, dass es vergessen wird oder fehlerbehaftet sein kann, und nicht zuletzt zeitintensiv und lästig für den Anwender ist[5]. Im Allgemeinen werden die speziellen Vorgehensweisen von dem ungelernten Benutzer soweit wie möglich versteckt, da solch eine fehlerbehaftete Konfiguration das gesamte Netzwerk in Gefahr bringen kann. Gleichzeitig nutzt man die schlechte Lokalisierbarkeit von mobilen Endgeräten zum eigenen Vorteil aus, um selbst einem speziell auf sich gerichteten Angriff entgehen zu können[6]. Durch das automatische Update und sofortige Integration neuer Mobile Nodes kann eine optimale Nutzung garantiert werden, so dass schließlich Intrusion Detection in dynamischer Umgebung möglich wird.

7.4.4 Vorgehensweise in Ad Hoc-Netzwerken

In Ad Hoc-Netzwerken ist es noch etwas komplizierter die Sicherheit zu waren. Möglicherweise könnten diese Lösungsansätze dazu beitragen, die auftretenden Probleme zu beseitigen.

Schichtübergreifende Intrusion Detection:

Auch bei Ad Hoc-Netzwerken können die Probleme in allen Schichten auftreten. Jede Schicht ist in mobiler Umgebung einfacher angreifbar. Die Sicherheit auf Mobile Nodes ist derjenigen im drahtgebundenen Netz noch deutlich unterlegen. Deshalb sollte ein Intrusion Detection System auf verschiedenen Layern arbeiten können. Bei einer Anomalie in einem der Layer wird dann den ID Sensoren, die auf den anderen Layern arbeiten, eine Warnung geschickt[6].

Lösungsansatz für Angriffe auf Ad Hoc-Netzwerke:

Die Dezentralisierung bedeutet für jedes Mobile Node einen hohen Grad an Verwundbarkeit. In Ad Hoc Netzwerken kommt hinzu, dass durch die Fremdartigkeit der Kommunikationspartner die Sicherheit nicht mehr gewährleistet werden kann. Deshalb sollte jedes MN erstmal grundsätzlich für sich selbst Intrusion Detection betreiben können[6]. Nun geht man einen Schritt weiter. Wenn die anderen MNs in der Lage sind mit dem eigenen MN Intrusion Detection zu betreiben, so sollte man zusätzlich zum eigenen Ergebnis die anderen MNs zu Rate ziehen[6]. Jedoch ist das nur als Ergänzung zu betrachten. Eine mögliche Ausführung sieht vor, in jedem MN einen Intrusion Detection System Agent installiert zu haben, und zunächst die Sammlung und Prüfung der Daten lokal, also nur für das eigene Mobile Node, durchzuführen. Weiterhin geht man so vor, dass man bei

einer auftretenden Anomalie mit den Nachbarn kommuniziert, und eine globale Prüfung der Daten anordnet[6]. Die zugesendeten Daten sollten dabei nicht verwendet werden, da eine absichtliche Fälschung oder bösartiger Code enthalten sein kann. Die Kommunikation zwischen sich und den Nachbarn wird durch eine so genannte 'secure communication engine' geführt, die den Datentransfer überwacht und versucht mehr Sicherheit zu gewährleisten[6]. Sie ist Bestandteil eines jeden IDS Agents. Da bei der Kommunikation mit näheren Nachbarn weniger fehlerhafte Datenpakete entstehen, da die Entfernung geringer ist und durchschnittlich weniger Störungen auftreten, erhalten ihre Ergebnisse eine höhere Priorität[6]. Wenn eine ausreichend eindeutige Bewertung durch die globale Untersuchung getätigt wurde, und deshalb davon ausgegangen werden kann, dass eine Anomalie tatsächlich vorliegt, kann eine globale Reaktion auf diesen Eingriff vorgenommen werden. Diese kann von jedem beliebigem Mobile Node getätigt werden, nicht zwangsweise durch das eigene. Die Maßnahme ist nun abhängig von der Art des Angriffs, der Art der Netzwerkprotokolle und wie hoch das Vertrauen in das Ergebnis ist, also wenn möglichst alle die Anomalie bestätigt haben[6]. Die Intrusion Response sieht vor, die Kommunikation zwischen den Teilnehmern neu zu initialisieren, in dem beispielsweise die Login-Keys neu angefordert werden[6]. Konnte der Eindringling identifiziert werden, wird eine Reorganisation der Netzwerkstruktur unter Ausschluss unerwünschter MNs eingeleitet. Problematisch ist außerdem die Datenkontrolle, da keine Router, Gateways oder andere Verkehrsknotenpunkte vorhanden sind, an denen man die zu prüfenden Daten auslesen kann. In einem gemeinsamen Medium ist das zwar nicht unbedingt erforderlich, da alle MNs alle Daten empfangen können, aber stattdessen entsteht in einem infrastrukturlosen Netzwerk das Problem der so genannten 'hidden stations', wobei die Erkennung eines belegten Kanals nicht in allen Situationen zuverlässig funktioniert. Hier kurz eine Erklärung des Problems. Eine Station kommuniziert mit einer zweiten, mit der eine weitere Station nun auch kommunizieren möchte. Wenn nun diese dritte Station die erste nicht erkennen kann, weil diese beispielsweise außerhalb der Reichweite liegt, geht sie davon aus, der Kanal wäre frei, und sendet die Datenpakete. Diese gehen jedoch verloren, da der Kanal in Wirklichkeit belegt ist. Dies kann natürlich dazu führen, dass die Daten nur unvollständig vorliegen. Hier tritt wieder das Problem auf, welche Algorithmen man für unvollständige Daten verwenden sollte.

7.5 Fazit

Diese Ausarbeitung diente im Rahmen eines Seminars dazu dem Interessierten näher zu bringen, worum es in der Intrusion Detection geht. Es wurden die Methoden der Intrusion Detection angerissen und auf die vielfältigen Varianten des Einsatzes von ID Systemen eingegangen. Entscheidend war besonders die Darlegung der Umsetzung von Intrusion Detection in die drahtlose Umgebung und welche Ansätze derzeit zu deren Verwirklichung existieren. Es wurde herausgestellt, dass ohne einige Änderungen vorzunehmen man Intrusion Detection also nicht von drahtgebundener Umgebung in drahtlose Umgebungen übernehmen kann. Hierbei wurde ein deutlicher Unterschied zwischen Infrastruktur-Netzwerken und Ad Hoc-Netzwerken gemacht. Beide Lösungen fordern die Mitarbeit der Kommunikationspartner des Netzwerkes, doch während man bei einem Infrastrukturnetzwerk

eine noch recht einfache und sichere Lösung hatte, ist die Lösung bei Ad Hoc-Netzwerken schwieriger zu realisieren und bietet einen nicht ganz so hohen Grad an Sicherheit. Zusammenfassend kann man sagen, dass für beides eine realistische Möglichkeit der Umsetzung existiert. Bei einem zunehmend bedrohlichen Angriffsszenario kann es sowohl für mobile, als auch für stationäre Systeme sinnvoll oder gar notwendig sein, Intrusion Detection Systeme einzusetzen.

Literaturverzeichnis

[1] S. Axelsson
Intrusion Detection Systems : A Survey and Taxonomy
March 2000

[2] J. von Helden, S. Karsch
Grundlagen, Forderungen und Marktübersicht für Intrusion Detection Systeme (IDS) und Intrusion Response Systeme (IRS)
http://www.bsi.de/literat/studien/ids/doc0000.htm
October 1998

[3] J. von Helden, S. Karsch
BSI-Leitfaden zur Einführung von Intrusion-Detection-Systemen
http://www.bsi.de/literat/studien/ids02/index.htm
October 2002

[4] Andreas Wespi, Giovanni Vigna, Luca Deri
Recent Advances in Intrusion Detection
December 2002

[5] Y. Zhang, W. Lee, Y. Huang
Intrusion Detection Techniques for Mobile Wireless Networks
January 2003

[6] Y. Zhang, W. Lee
Intrusion Detection in Wireless Ad Hoc Networks

Kapitel 8

Software Environments for Mobile Devices

Michael Böhm

Der Markt für mobile Geräte hat in den letzten Jahren immens an Bedeutung gewonnen. Wegen der begrenzten Resourcen derartiger Systeme werden auch besondere Anforderungen an entsprechende Betriebssysteme gestellt. Die gängisten Betriebssysteme im Bereich von mobilen Systemen sind Symbian, die Microsoft Familie und natürlich Linux. Symbian ist im Bereich von Handys am weitesten verbreitet und bietet den großen Vorteil, dass alle Partner des Konsortiums den Sourcecode abändern und an ihre Bedürfnisse anpassen können. Microsoft macht die Hersteller zu reinen Hardwarelieferanten, hat aber den Vorteil, dass durch die weite Verbreitung von Microsoft Betriebssystemen auf PCs das Abgleichen von Daten mit dem Pc erleichtert wird. Letztendlich bleibt noch Linux. Linux bietet den Vorteil, dass es sehr Zuverlässig und stabil arbeitet. Leider ist es im Bereich von handys noch immer ein Nischenprodukt. Auch an die Programmierumgebung für mobile Geräte werden spezielle Anforderungen gestellt. Auch hier spielen die begrenzten Resourcen eine große Rolle. Java stellt dafür J2ME zur verfügung, eine eingeschränkte Version für derartige Systeme. Da das API, das mit dieser Konfiguration ausgeliefert wird, dem Entwickler nur einige Grundfunktionalitäten zur Verfügung stellt, bedarf es einer Erweiterung dieses API's mittels eines zusätzlichen Profils. Das Profil wird in diesem Einsatzgebiete mit "Mobile Information Device Profile" (MIDP) bezeichnet.

Inhaltsverzeichnis

8.1 Einleitung 167
8.2 Anforderungen an Betriebssysteme für mobile Systeme . . . 167
8.2.1 Stabilität und Ausfallsicherheit 168
8.2.2 Speicherbedarf 168
8.2.3 Powermanagement 169
8.2.4 Echtzeiteigenschaften 169
8.3 Die gängigsten Betriebssysteme - Überblick 169
8.3.1 Symbian 169
8.3.2 Microsoft Familie 171
8.3.3 Linux 172
8.4 Aufbau am Beispiel von Symbian 175
8.4.1 Hardware-Sicht 175
8.4.2 Application-Sicht 179
8.5 Bewertung 184
8.5.1 Kernel 185
8.5.2 Portabilität 186
8.5.3 Skalierbarkeit 186
8.5.4 Performance 186
8.6 Fazit 186
8.7 Java 2 Micro Edition (J2ME) - Java Umgebung für Embedded Anwendungen 187
8.7.1 Einordnung und Zielsetzung von MIDP 187
8.7.2 Architektur 187
8.7.3 Voraussetzungen 188
8.8 Ausblick 189

8.1 Einleitung

Diese Arbeit beschäftigt sich mit den gängigsten Betriebssystemen für mobile Systeme und soll einen kurzen Einblick in die Programmierplattform J2ME ermöglichen. Zunächst soll auf allgemeine Anforderungen für derartige Betriebssysteme eingegangen werden, bevor die gängigsten Betriebssysteme im einzelnen näher betrachtet werden. Nachdem Symbian in Europa klar den Markt beherrscht, soll an diesem Beispiel der Aufbau eines Betriebssystems für mobile Systeme erläutert werden. Danach werden die hier vorgestellten Systeme anhand bestimmter Kriterien bewertet. Zuletzt beschäftigt sich diese Arbeit noch mit J2ME und zeigt auf, was für die Zukunft auf diesem Markt noch zu erwarten ist.
Was sind mobile Systeme? Mobile Systeme sind Rechensysteme, die mitgeführt werden können (Smartphones, Handys, Palmtops, etc.), oder die sich selbst bewegen (Auto, Fugzeug, etc.). Letztere spielen aber für diese Arbeit eine untergeordnete Rolle. Obwohl die Grenzen zwischen Betriebssystemen für mobile und stationäre Systeme immer mehr verschwimmen, lassen sie sich in manchen Punkten klar voneinander abgrenzen.
Was ist ein Betriebssystem? Ein Betriebssystem ist die Organisations- beziehungsweise Verwaltungseinheit innerhalb von Rechensystemen. Die primäre Aufgabe ist die Verwaltung der zur Verfügung stehenden Betriebsmittel. Für den Benutzer erheblich ist allerdings der Zugang zum System über eine virtuelle Maschine, die er leichter handhaben kann als die Maschine selbst.
Was ist ein embedded System? Betriebssysteme, die keinen direkten Zugriff auf sich selbst erlauben, nennt man embedded Systems (engl.: eingelassen, verborgen). Sie kommen im Bereich der Echtzeit-Systeme besonders zum Tragen, auch wenn das wiederum nur einen Teil darstellt. Weit über 90% aller Systeme werden auf embedded Systems zurückgeführt, was auch die Grundlage für diesen hart umkämpften Markt ist. Es dürfte schnell einsehbar sein, dass der Markt für gängige Betriebssysteme in diesem Zusammenhang nur ein "kleiner Fisch" ist. [9]
Man kann schon an dieser Stelle einige Probleme erkennen, die sich im Zusammenhang mit mobilen Systemen und deren Eigenschaften ergeben. Es ist verständlich, dass man an Betriebssysteme für mobile Systeme, die keinen Einfluss vom Endverbraucher erlauben und eingebettet in ihre Umgebung ohne Reboot oder dauerhafte Stromversorgung auskommen müssen, ganz besondere Anforderungen stellt. Die Tatsache, dass sich mobile Systeme immer größerer Beliebtheit erfreuen, macht deutlich, dass in dieser Richtung noch einiges geschehen muss und wird.

8.2 Anforderungen an Betriebssysteme für mobile Systeme

Mobile Systeme jeglicher Art verdanken ihren hohen Nutzwert ihrer Mobilität. Seien es mobile Telefone, PDAs oder auch solche mobile Systeme in modernen Autos, sie alle würden ihren Sinn verlieren, wenn sie stationär gebunden wären. Um Anforderungen festhalten zu können, muss man im Allgemeinen daher, gebunden an die Fragen "Wie hoch soll die Mobiliät sein?" und "In was für einem Endprodukt soll das Betriebssystem eingesetzt werden?", grundlegend zwischen folgenden Gesichtspunkten unterscheiden:

- datenspezifisch
  Wie wird die Speicherung realisiert?
  Wie hoch darf/soll die externe/interne Kommunikation sein?
  Welche Transfergeschwindigkeiten sind erwünscht?
- versorgungsspezifisch
  Wie wird die Stromversorgung geregelt?
  Wie lange muss das System ohne Zugang zu einem festen
  Stromnetz auskommen können?
  Wie sind Updates oder Anwendungen Dritter zugänglich?
- sicherheitsspezifisch
  Welche Angriffspunkte birgt die Mobilität?
  Inwieweit schränkt eine höhere Sicherheit andere Gebiete ein?
  Wie schützt sich das System gegenüber Ausfällen oder Eingriffen durch
  Anwendungen eines Drittherstellers?
- produktspezifisch
  Ist minimaler Platz erforderlich?
  Sind besondere Anwendungen verlangt/ausgeschlossen?
  Wie kann ich das Betriebssystem erweitern, verändern oder anpassen?

Natürlich könnte man noch einige Fragen ergänzen, doch sollen die hier aufgeführten Fragen für diese Arbeit ausreichen. Wie in vielen anderen Bereichen auch, sind Überschneidungen der einzelnen Fragen nicht zu vermeiden. Mit Hilfe der gestellten Fragen soll nun versucht werden, die Anforderungen an Betriebssysteme für mobile Systeme zu erarbeiten und festzuhalten. Auch wenn die einzelnen Hersteller verschiedene Schwerpunkte setzen und Probleme auf verschiedene Art und Weise zu lösen versuchen, stehen sie doch den gleichen Problemstellungen gegenüber.

8.2.1 Stabilität und Ausfallsicherheit

Ausfälle und fehlende Stabilität sind bei PCs keine Seltenheit. Haben diese Probleme bei den herkömmlichen PCs keine schwerwiegenden Folgen, ist das bei enbedded Systems anders. Im Mobilfunkbereich ergeben sich zwar keine schwerwiegenden Folgen, doch wird es der Verbraucher nicht hinnehmen, wenn starke Schwankungen in der Übertragungsqualität auftreten, oder die Verbindung öfter zusammenbricht. In Anbetracht der Einsatzgebiete sind diese Anforderungen für embedded Systems unabdingbar. Will man schon hier einen kurzen Vergleich mit gängigen PC-Betriebssystemen anstellen, dann grenzt ein solches Unterfangen schon fast an eine unmögliche Aufgabe. [8]

8.2.2 Speicherbedarf

Der minimale Speicherbedarf beziehungsweise die Größe des Systems ist stark abhängig vom Einsatzgebiet. Muss man beispielsweise ein Betriebssystem für ein Mobiltelefon entwerfen, dann ist zum einen nicht viel Platz für ein umfangreiches System und zum anderen

sind die Kundenwünsche stark “eingeschränkt“. Neben Telefonie sollen bestimmte Applikationen im Umfang des Angebotes zur Verfügung stehen, aber die Bootzeit (Starten des Betriebssystems) darf sich dabei, abgesehen einer gewissen Toleranz, nicht verändern. Entwirft man allerdings ein Betriebssystem im Bereich einer Auto- oder Jetsteuerung, dann ist gegebenenfalls der Platz Nebensache und die Ausfallsicherheit bzw. die Sicherheitssubroutinen stehen im wesentlichen Vordergrund. Die Folge ist klar ersichtlich: Spezialisierung. [7]

8.2.3 Powermanagement

Einige mobile Systeme müssen eine lange Zeit (z.B. mehrere Tage) ohne Stromversorgung auskommen oder im Falle eines Mobiltelefons noch aktiv sein, obwohl das Gerät abgeschaltet wurde, damit beispielsweise der Benutzer mit einem Alarm an einen Termin erinnert werden kann. Ein jeweiliges Betriebssystem darf dem nicht hinderlich im Weg stehen, sondern muss mit einer intelligenten Ressourcenverteilung und -nutzung das Management verbessern und dafür sorgen, dass es genügsam mit vorhandenen Betriebsmitteln umgeht.

8.2.4 Echtzeiteigenschaften

Es wäre wünschenswert, wenn man Echtzeit-Verhalten in manchen Bereichen erzielen könnte, doch sind diese Eigenschaften nicht unbedingt notwendig. Allerdings ist klar, dass der User gewisse Dinge hinsichtlich “Echtzeit“ erwartet. Das System sollte also durchaus in der Lage sein, bestimmte Aufgaben in bestimmten Zeitintervallen zu erledigen. So erwartet man z.B., dass bei einem eingehenden Anruf die momentane Anwendung unterbrochen wird, um den Anruf entgegenzunehmen. Symbian hat dieses Problem aufgegriffen und mit der neusten Version Symbian OS V8.0 versucht zu lösen. Laut Microsoft ist ein echtzeitfähiges Betriebssystem durchaus realisierbar [4]

8.3 Die gängigsten Betriebssysteme - Überblick

Die zwei wichtigsten Hersteller für derartige Betriebssystem sind Symbian und Microsoft. Beide konkurrieren auf dem Markt um die Vorherrschaft. Wobei der europäische Markt momentan fest in Händen von Symbian ist. Aber Microsoft befindet sich auf dem Vormarsch. Als drittes wichtiges Betriebssystem müssen wir noch Linux betrachten, auch wenn dieses Betriebssystem einige Vorteile bietet, ist es bei weitem nicht so weit verbreitet, wie Symbian OS oder die Betriebssysteme der Microsoft Familie.

8.3.1 Symbian

Symbian wurde vor sieben Jahren von Nokia und anderen Branchengrößen gegründet, um ein Betriebssystem für Smartphones zu entwickeln. Das Joint Venture war dabei höchst

erfolgreich, rund 94 Prozent aller Smartphones, die etwa im zweiten Quartal in Europa, dem Nahen Osten und Afrika verkauft worden sind, laufen mit Symbian-Software. Microsoft blieb mit seinem Handy-Betriebssystem dagegen in Europa bisher eher erfolglos. Allerdings hatte sich Motorola im August 2002 von Symbian zurückgezogen und war eine Allianz mit Microsoft eingegangen.

Trotz dieser neuen Allianz ist Symbian-Chef Levin auch für die Zukunft optimistisch: Microsoft werde sich am Mobiltelefon nicht durchsetzen, weil es die Handyproduzenten zu reinen Hardwarelieferanten mache und deshalb die Gewinnmöglichkeiten reduziere. Symbian dagegen liefere nicht, wie es der Redmonder Konzern praktiziere, eine Software an die Kunden, mit der sie ohne Wenn und Aber leben müssten. "Jeder Lizenznehmer kann die Software individuell anpassen. Jedes Symbian-Handy von Nokia oder Sony-Ericsson ist mit anderen Diensten ausgestattet. Trotzdem "sprechen" beide Handys miteinander", sagte Levin.

Die Firma Symbian stellt Betriebssysteme für mobile Telefone her. Dabei legt sie ihr Marktsegment sowohl auf die sogenannten Smartphones (Telefone mit PDA-Fähigkeiten) als auch auf die Communicators (PDAs mit Telefonie). Laut diverser Firmen kann man erkennen, dass sich dieser Markt überschlägt und fast monatlich neue Modelle oder Anwendungen für den Endbenutzer zugänglich werden. Innerhalb dieser rasanten Entwicklung scheinen Stabilität und Ausfallsicherheit fast sekundär, wenn man Probleme moderner Handys betrachtet. Grundlegend gibt es laut Symbian White Papers ganz andere Eigenschaften, die hier eine primäre Rolle spielen. Es sollen nun einige dieser besonderen Anforderungen aufgelistet werden.

- **Funktionalität**
  Gemäß der rasanten Entwicklung muss ein Betriebssystem zukunftsorientiert ausgelegt sein. Es darf nicht schon mit der nächsten Hardwareentwicklung veraltet sein. Man unterscheidet zwischen Hilfsfunktionalität (Anwendungen wie Rechner, Musikspieler, ein Spiel) und Integritätsfunktionalität (z.B. verbesserte Datenbanken, intelligentere Speicherung von Daten). In diesem Zusammenhang spielt auch die Anpassungsfähigkeit eine Rolle. Denn nur das Betriebssystem, das auch die nächste Generation Mobiltelefone unterstützt, kann sich auch langfristig auf dem Markt behaupten.
- **Spezielle Hardware**
  Bei der Größe des Endprodukts und der hohen Leistungsfähigkeit eines dazugehörigen Betriebssystems müssen spezielle Bausteine (CPU, Speicher, etc.) integriert werden. [8]
- **Offenes System**
  Ausgelegt als ein offenes Betriebssystem, das sich als Standard für den Mobiltelefonmarkt einbürgern sollte, muss sich das System erweitern und verändern lassen. Auch das stellt besondere Anforderungen an das System, da es in seinen Grundfesten bei Änderungen nicht einfach "auseinanderbrechen" darf.
- **Ausfallsicherheit und Stabilität**
  Der Endbenutzer sieht es als selbstverständlich Software oder Anwendungen für seinen Organizer zu laden. Um die Integrität des Systems zu wahren, muss der Betriebssystemkern stabil und unempfänglich gegen Ausfälle sein.

- **Effiziente Nutzung der begrenzten Ressourcen**

  Man muss sich vor Augen führen, dass physikalische Netze immer schneller sein werden als die drahtlose Übertragung. Der Versuch in direkte Konkurrenz zu treten wäre im vornherein zum Scheitern verurteilt. Daher muss man gerade in diesem Bereich darauf zurückgreifen was man hat. Wenn man also die globale Geschwindigkeit von Übertragungsraten selbst nicht erhöhen kann, dann muss dafür gesorgt werden, dass die Übertragung selbst effizient und schnell abgewickelt werden kann. Das System muss in Echtzeit in der Lage sein den Nutzer über den aktuellen Status zu informieren und mittels geeigneter Protokolle jede Phase von Erreichbarkeit sinnvoll auszunutzen.

8.3.2 Microsoft Familie

Microsoft hat selbstverständlich auch in diesem Markt ein Interesse Präsenz zu zeigen. Hauptaugenmerk der Firma war es allerdings besonders flexible und skalierbare Plattformen zu entwickeln, weniger sich auf ein spezielles Marktsegment festzulegen. Das neuste Betriebssystem von Microsoft ist Windows Mobile 2003. Zum Verständnis, was Windows Mobile 2003 ist, muss etwas ausgeholt werden, denn Microsoft verfolgt eine Produktstrategie, die von der anderer Entwickler von Mobilcomputer-Betriebssystemen abweicht. Die wichtigste Konkurrenz, Symbian mit SymbianOS, bedient nur jeweils einen stark abgegrenzten Markt, also Handhelds, beziehungsweise Smartphones, während Microsoft in jedem Lebensbereich präsent sein will. Microsoft hat daher seit 1996 an Windows CE gearbeitet, welches in angepasster Form in Unterhaltungselektronik, Handhelds, aber auch Smartphones zum Einsatz kommt. Das in den stiftbedienten PDAs, den Pocket PCs, verwendete Betriebssystem basiert auf Windows CE, wurde aber um spezielle Anwendungen wie Textverarbeitung, Tabellenkalkulation, Media Player und anderes ergänzt, die bei anderen Windows CE-basierten Produkten in dieser Form nicht zu finden sind. Tatsächlich basierte das Betriebssystem weiterhin auf Windows CE 3.0, bekam aber zusätzliche Anwendungen zur Seite. Irritierend ist, dass Microsoft seitdem auch von “Windows powered“-Geräten spricht, obwohl sich die Kompatibilität mit Desktop-Windows auf den Datenaustausch beschränkt. Anwendungen vom Windows-PC laufen dagegen nicht auf dem PDA. Das am 23. Juni 2003 vorgestellte, neueste Betriebssystem heißt jetzt “Mobile Windows 2003“ und basiert nicht mehr auf den Betriebssystemkern “Windows CE“, sondern auf den Nachfolger “Windows CE.NET“. Die meisten Änderungen sind unter der Oberfläche zu finden und betreffen die drahtlose Kommunikation. Als Pocket PC 2002 auf den Markt kam, spielten Bluetooth und WLAN noch keine Rolle und wurden deshalb auch nicht von Microsoft im Betriebssystem berücksichtigt. Die Folge war, dass die PDA-Anbieter und Steckkarten-Hersteller eigene Treiber entwickelten und mitlieferten. Für den Kunden hatte dies den Nachteil, dass sie mit zum Teil mangelhaften Treibern der Hersteller herumplagen mussten. Auch der Verbindungsaufbau mit VPNs (Virtual Private Networks) ist erleichtert worden und der neue Internetprotokoll-Standard Ipv6 und IPSec/L2TP wurden integriert.

Veränderungen von Windows Mobile 2003 gegenüber Microsoft Windows CE:

- Drahtlose Kommunikation per Bluetooth und WLAN (Wireless LAN) wird direkt vom Windows Mobile 2003-Betriebssystem unterstützt.
- Komplett überarbeitet hat Microsoft die Verwaltung der Modem- und Netzwerkverbindungen.
- Nützliche Features erleichtern in der Posteingang-Anwendung die Email-Verarbeitung.
- Die Verbesserungen am integrierten Webbrowser verbergen sich unter der Oberfläche.
- Zahlreiche Verbesserungen im Multimedia-Bereich.

Diese Betriebssysteme versteiften sich eher auf die PC-ähnlichen embedded Systems. Also all die Systeme, die auf ein umfangreiches Betriebssystem zurückgreifen konnten, da es keine Unterbringungsprobleme gab. Zur MS Embedded Serie gehören Windows CE .NET und Microsoft Mobile 2003.

Auch hier entstanden weitere Anforderungskriterien, die sich durch die Entwicklung, die Studien und die Erfahrung am Markt ergeben hatten. Es soll hier nur auf die Anforderungen [4] eingegangen werden, die sich im Rahmen der Arbeiten an Windows Mobile 2003 zusätzlich ergeben haben, weil Windows CE in dem bisherigen Rahmen schon erfasst wurde.

- **Skalierbarkeit**
  Keinen direkten Einschränkungen unterworfen, versuchte man näher an der ursprünglichen PC-Architektur zu entwerfen. Das Betriebssystem musste in dem Rahmen eine reiche Funktionalität aufweisen, aber vor allem eine Möglichkeit des modularen Aufbaus bieten. Bei der Entwicklung von Mobile 2003 legte Microsoft besonders großen Wert auf diese Eigenschaft.
- **Flexibilität**
  Das System muss besonders anpassungsfähig sein, wenn es auf keinen Markt speziell ausgerichtet ist. Theoretisch müsste es für jedwede denkbare mobile Anwendung zugänglich sein. Ein solches Unterfangen birgt eine große Komplexität für den Entwickler.
- **umfangreiche Sicherheit**
  Mit einer hohen Skalierbarkeit und einer PC-ähnlichen Struktur “gewinnt“ man auch die Sicherheitsnachteile. Dahingehend muss das Betriebssystem wieder umfangreich geschützt werden.

8.3.3 Linux

Zunächst stellt sich die Frage, warum gerade Linux dazuprädestiniert sein soll in eingebetteten Systemen eingesetztzu werden. Natürlich spielen die Kosten eine große Rolle. Da embedded systems in sehr großen Stückzahlen produziert werden, sind schon geringe Lizenzkosten ein großer Finanzfaktor für den Hersteller. Ein weiterer Punkt ist die Zuverlässigkeit des Systems, da auch hier Ausfallzeiten schnell sehr teuer werden können. Die

Systeme sind viele Jahre im Einsatz und auftretende Bugs sollten ohne größere Schwierigkeiten beseitigt werden können. Da ist es natürlich besonders vorteilhaft, wenn der gesamte Source zu Verfügung steht. Linux ist wohl im Moment eines der stabilsten und zuverlässigsten Betriebssysteme auf dem Markt. Durch die open source Verbreitung ist es auch besonders flexibel, wenn es darum geht, ein bestehendes System auf ein anderes zu portieren. Es würde viel Zeit und Geld kosten, sich sämtliche Funktionen mühsam zu erschließen und selbst zu programmieren. Weil embedded systems meist klein und billig sein sollen, werden häufig nicht mehr ganz aktuelle PC-Komponenten zu ihrer Herstellung verwendet. Das Betriebssystem sollte dann alles aus dieser "low-power" Hardware herausholen. Auch hier sticht Linux, im Gegensatz zu anderen Betriebsystemen, durch seinen geringen Sourcenverbrauch hervor. Selbst Linux ist nicht auf jeder Hardware lauffähig. Gewisse Mindestanforderungen sollten schon gegeben sein, damit das System einwandfrei funktioniert. Man spricht in der Regel von einem Bedarf von 2 MB RAM, 2 MB Flash und einem 32 Bit Mikroprozessor. Natürlich gibt es keine Regel ohne Ausnahme, es gibt bereits Linux Kernel die in 80 KB Speicher Platz finden und damit theoretisch nur 256 KB RAM benötigen würden. Linux kann selbstverständlich selbst bei sehr kleinen Zielsystemen als Entwicklungs-umgebung eingesetzt werden, hierfür stehen sehr viele, frei verfügbare, Softwarewerkzeuge zur Verfügung. Oft kann sogar das Testen der Software auf dem Zielsystem ausbleiben, da Linux die Hardware sehr gut emulieren kann. Das spart Entwicklungszeit und damit viel Geld. Eine weitere Schwäche des Standardlinux ist seine nicht vorhandene harte Echtzeitfähigkeit. Zwar übertrifft die implementierte weiche Echtzeitunterstützung die anderer Betriebssysteme bei Weitem, aber für den industriellen Einsatz müssen Interrupt Antwortzeiten von wenigen Mikrosekunden gewährleistet werden. Auch für dieses Problem gibt es unter Linux eine Lösung, die sog. Betriebssystemerweiterungen. Hier hat Realtime-Linux (RT-Linux) bereits seine Industrietauglichkeit bewiesen. Linux stellt auch die Ausgangslage für Dutzende Embedded Projekte dar. So kann es auch als Betriebssystem für mobile Endprodukte genutzt werden. Besonders betont werden hier die folgenden Eigenschaften [3]:

- **Stabilität**

  Der Linux Kernel weist bereits eine hohe Stabilität und Ausfallsicherheit auf und ist damit für mobile Anwendungen nutzbar zu machen. Das System kann ohne weiteres mehrere Wochen und Monate ohne Probleme laufen und wird daher auch in vielen Bereichen (z.B.: Webserver, etc.) erfolgreich eingesetzt. Abstürze sind in den meisten Fällen auf Hardware-Fehler oder die Stromversorgung zurückzuführen. Hinsichtlich Stabilität ist Linux den anderen vorgestellten Betriebssystemen überlegen.

- **Performance**

  Das Betriebssystem kann genügsam mit Ressourcen und Speicher umgehen. Ein effizientes Management sorgt für den sinnvollen Einsatz von CPU Power, Arbeitsspeicher und Festplattenspeicher. Dennoch ist dieser 'Wert' nur schlecht zu messen. Eine schlechte Performance begründet sich in einem instabilen Kern und Anwendungen, die auf einem solchen System ausgeführt werden, können abstürzen, obwohl die Software an sich einwandfrei funktioniert. Um die Performance zu testen, kann man Benchmarks einsetzen. Diese Software versucht die Leistung eines Systems mittels zahlreicher Testdurchläufe unter verschiedenen Bedingungen zu ermitteln. [6] Dabei entsteht ein Richtwert, den man mit anderen Systemen vergleichen kann. Dieser Wert

ist stark abhängig von der Zielanwendung und dementsprechend von der Zusammenstellung des Betriebssystems.

- **Skalierbarkeit**

  Linux wird bereits auf unterschiedlichsten Plattformen verwendet. Es ist in der Lage sich auch extremen Bedingungen anzupassen, da es dank der Tatsache, dass es ein offenes Betriebssystem ist, in verschiedenen Bereichen angepasst und weiterentwickelt wurde. Der eigentliche Linux-Kernel ist allerdings über 1.5 MByte groß. Mit weiteren Anwendungen gelangt man schnell auf mehrere Megabyte. Damit das System für mobile Systeme genutzt werden kann, muss man sich die Modularität nutzbar machen und den Kernel feiner konfigurieren. Dadurch kann man Werte im Bereich von 300 Kilobyte erreichen. Und auch wenn der Kernel mit TCP/IP um die 500 Kilobyte groß ist, so ist der memory footprint immer noch akzeptabel.

- **Portabilität (Offenes System)**

  Linux hat bereits eine breite Nutzung auf unterschiedlichen Plattformen erreicht, so dass man in der Entwicklung des Betriebssystems darauf fokussiert war, die Kompatibiltät zu möglichst vielen Plattformen zu forcieren. Dadurch wird durch eine neue Hardware nicht zwangsweigerlich ein neues Betriebssystem benötigt. Zudem fördert die Tatsache, das Linux ein offenes System ist, die Entwicklung durch eine breite Front an mitwirkenden Herstellern und auch Benutzern.

- **Echtzeit-Verhalten**

  Linux ist von sich aus nicht in der Lage in Echtzeit zu reagieren. Prozesse müssen im Queue warten, damit andere Prozesse beendet werden können oder wiederum auf den Queue gelegt werden. Die Lösung dieses Problems liegt in dem Versuch einen RT-Kernel zu erstellen, der genau dieses Echtzeit-Verhalten an den Tag legt. LynxOS ist ein Linux-basiertes Betriebssystem, das Echtzeiteigenschaften nach dem Standard DO-178B aufweist. Dieser Standard ist in den Vereinigten Staaten die Voraussetzung für sämtliche Software im Bereich der Flugtechnik und ein embedded System muss diese Mindestanforderungen erfüllen. Die Realisierung wird beispielsweise durch Partitionierung und eine abstrakte Zuordnung durchgeführt. Dabei werden den Anwendungen einzelne Bereiche zugeordnet, damit sie praktisch in einer eigenen Betriebssystemumgebung operieren können und theoretisch eine 'eigene CPU' beanspruchen.

Zudem wird der Linux Kernel fortwährend weiterentwickelt und dieser Prozess vollzieht sich sehr schnell. Die neuesten Kernel Designs beinhalten wie schon erwähnt Modularität. Das bedeutet, dass der Kern selbst sehr klein gehalten wurde und die Erweiterungen komponentenweise ergänzt werden. Gerade in diesem Bereich herrscht zwischen den Linux- und den Windows- Anhängern ein reger Diskussionsbedarf. Und auch wenn durch einige neuere Abhandlungen die Windows Embedded Serie von verschiedenen Firmen mit Linux Embedded Systemen im Vergleich als schlechter dargestellt wird, so kann Windows Mobile 2003 durchaus als argumentative Grundlage für die Skalierbarkeit eines Systems dienen.

8.4 Aufbau am Beispiel von Symbian

In diesem Teil soll der Aufbau eines Betriebssystems für mobile Systeme erläutert werden. Dabei wird das Betriebssystem von zwei Seiten betrachtet. Die eine, die Hardware-Sicht, beschäftigt sich ausführlich mit der benötigten Hardware zum einen und der programmtechnischen Lösung zum anderen. Eine zweite Sicht wirft die der Applications auf. Dadurch erhält man einen besseren Zugang zu den einzelnen Anwendungen und ihre Abhängigkeiten. Weiterhin soll das Zusammenspiel der einzelnen Teile erläutert und der modulare Aufbau eines Betriebssystems erklärt werden. Es wird auf die Skalierbarkeit und auch die unterschiedliche Bandbreite eines solchen Systems hingewiesen.

8.4.1 Hardware-Sicht

In diesem Kapitel soll auf drei Aspekte der Hardware-Sicht näher eingegangen werden. Dies sind im Einzelnen

- der Hardware-Aufbau und damit die Positionsbestimmung der Einzelteile
- der Aufbau des Betriebssystems und damit die Funktionsweise der Einzelteile
- und zuletzt der Betriebssystemkern selbst, der für die Kontrolle und Steuerung der Einzelteile zuständig ist

Hardware-Aufbau

Die 3-Layer-Architektur setzt drei, standardisierte Ebenen als Grundlage voraus (siehe Abbildung 1). Dadurch wird das Zusammenspiel einzelner Komponenten erleichtert und der Befehlssatz des Betriebssystems optimiert.

- **First Layer - CPU Core**

  Auf der ersten Ebene, dem Kern des Systems, befindet sich eine schnelle und kostengünstige CPU, die zudem nur wenig Energie verbraucht. Neben dem zentralen Prozessor befinden sich hier auch solche grundlegenden Einrichtungen wie die Memory Management Unit (MMU) und die Caches. Symbian OS setzt eine integrierte MMU und einen Cache voraus, um privilege levels zu ermöglichen. Mit Interrupts und Exceptions wird somit eine Prioritätencodierung zugänglich, die in der Form benötigt wird, um den schnellen und problemlosen Ablauf von Prozessen zu gewährleisten. Zur Erläuterung der einzelnen Begrifflichkeiten sei hier auf Fachliteratur verwiesen. [5]

  Main Memory Unit

  An dieser Stelle soll das Konzept der Virtual Machine Environment (VME) geschildert werden. Es dient dazu, um Prozesse mit Hilfe der MMU im virtuellen Adressraum zu verschieben (siehe Abbildung 2). Dadurch kann schon im voraus Speicherbedarf berechnet werden und man spart sich wiederum Zeit und Belastung der CPU, in dem man die Daten nach einer Löschung erneut laden müsste[8].

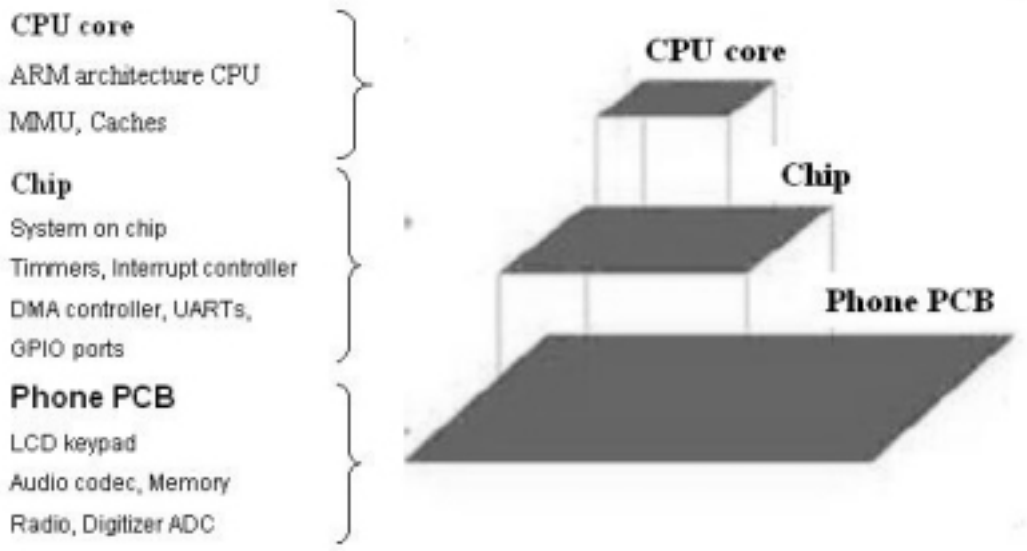

Abbildung 8.1: Die 3-Layer-Architektur. [8]

Cache und Speicher

Eine weitere wichtige Frage für die Unterbringung von Daten und des Betriebssystems selbst ist die nach dem Speicher. Nur ein geringer Teil des Speichers ist on-chip verfügbar. Der größte Teil liegt außerhalb des Hardwarekerns und wird off-chip-memory genannt. Dieser hat vor allem die Aufgabe das Symbian OS image (siehe weiter unten) zu speichern, damit die CPU darauf ständigen Zugriff hat. Weiterhin werden alle anderen wichtigen Daten (User-, Prozessdaten) hier gespeichert. In Bezug auf verschiedene Speicherbausteine, deren Vor- und Nachteile sei auch an dieser Stelle wieder an geeignete Fachliteratur von [5] verwiesen. Es soll an dieser Stelle genügen, dass diese Problematik in den meisten Fällen durch Flash Memory gelöst wird. Dieser ist zwar sehr teuer, dafür aber auch schnell und wiederbeschreibbar.

- **Second Layer - SoC**
  Die CPU wird wiederum auf einen Chip integriert und stellt mit ihm zusammen die zweite Ebene dar. Dieser integrale Bestandteil trägt den Namen system-on-chip (SoC) und beinhaltet sämtliche lebenswichtigen Peripherigeräte für das Betriebssystem. So zum Beispiel auch die Timer, den DMA Controller und vieles mehr.

- **Third Layer - PCB**
  Auf der dritten Ebene werden alle restlichen Komponenten, die das eigentliche Endprodukt ausmachen, angelegt. Zusammengefasst trägt sie den Namen Printed Circuit Board (PCB).

Aufbau des Betriebssystem - Das Symbian OS image

Das eigentliche Betriebssystem wird im Symbian OS image gespeichert. Es besteht aus sogenannten dynamic linked libraries (DLL) und vielen, weiteren Dateitypen. Weitere Dateien sind zum Beispiel Fonts, Bitmaps und andere. Die DLLs bilden die Kernsubstanz

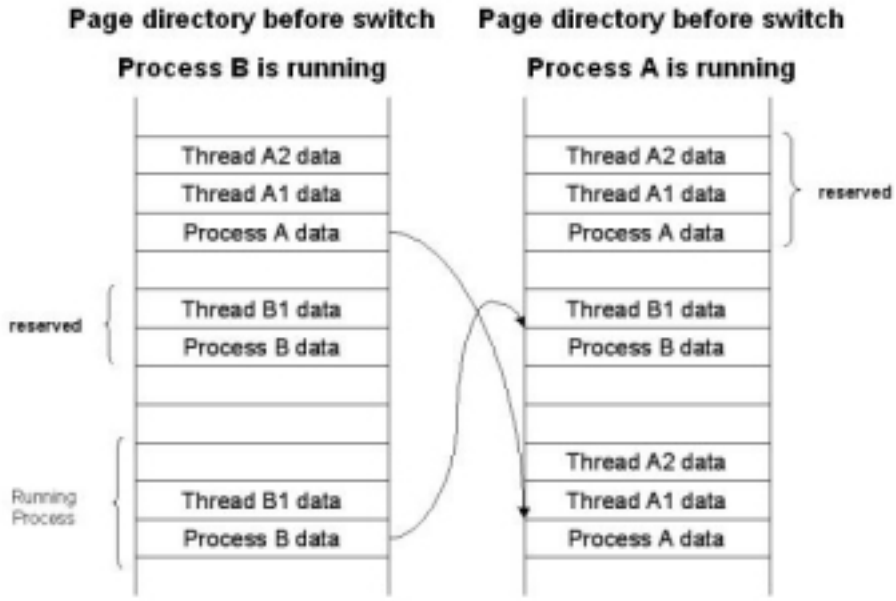

Abbildung 8.2: Die Virtual Machine Environment. [8]

des Systems und sind der wichtigste Schritt zur Lösung des Problems der Geschwindigkeit und Flexibilität eines Systems. Innerhalb dieser Dateien befinden sich wichtige Codesegmente, die sich durch ausführbare Dateien aufrufen lassen. Durch Wiederverwendung und häufige Kopplung kann so Speicherplatz und Ladezeit gespart werden. Der CPU führt das Betriebssystem direkt an seinem Speicherort aus. Falls das aber zu langsam sein sollte, kann man bestimmte Bibliotheken in den Random Access Memory kopieren. Dort können sie dann genutzt und nötigenfalls mit mehreren Prozessen verbunden werden. Mittels eines speziellen Schlüsselverfahrens können zahlreiche DLLs mit Anwendungen verbunden und gleichzeitig die Größe der Libraries minimiert werden. [8] Im Falle von Symbian OS wird ein spezieller DLL Typ häufig eingesetzt- polymorphic DLL. Diese Bibliotheken sind mit der Nutzbarmachung der objekt-orientierten Programmierung entstanden. An der ersten Position in der DLL steht eine Methode zum Aufruf der bestimmten Klasse. Beispielsweise kann man in der application DLL mit NewApplication() eine Instanz einer Anwendung aufrufen. [8] Im Sinne der Polymorphie hängt das Ergebnis des Aufrufs einer Methode von der Klasse ab, die die Methode nutzen will. Beispielsweise führt zeichne() bei dem Aufruf durch die Klassen Dreieck und Quadrat zu unterschiedlichen Ergebnissen, obwohl in beiden Fällen auf dieselbe Operation zugegriffen wurde.

Steuerung und Kontrolle - Der Symbian Kernel

Der Symbian Kernel (Betriebssystemkern) ist weniger als 200 KB groß. Im Verhältnis zu anderen Systemen ist das ein sehr guter Wert. Trotz des kompakten Baus unterstützt er beispielsweise Multitasking und ist fast vollständig von peripheren Geräten unabhängig. Die Hardwareunterstützung wird über DLLs geregelt, die verschiedene Komponenten

ansprechen können. Dabei wird eine dynamische An- und Abkopplung dieser Elemente unterstützt und fördert somit die Kompatibilität. Dabei wird der benötigte Code, um eine bestimmte Hardware anzusprechen, in abgegrenzte Bibliotheken gespeichert. Das setzt unter anderem eine intelligente Programmierung und eine zukunftsorientierte Planung voraus. Auf die entsprechenden Hardwarebereiche kann nur im privileged mode zugegriffen werden, was den Missbrauch durch andere Geräte verhindert. Da der Kernel immer in diesem Modus ausgeführt wird, hat er sämtlichen Zugriff zu allen Bereichen. Anwendungen greifen über Schnittstellen auf Kernel Dienste zu. Dies wird durch APIs (Application Programming Interface) realisiert und durch die Benutzerbibliothek unterstützt. Sämtliche Anwendungen laufen im unprivileged mode und können daher auch auf keinerlei Hardwarebereiche zugreifen. Benötigen sie diesen Zugang, dann muss diese Anwendung entweder die Berechtigung durch Mode-Switching erhalten oder über den Kernel Server vermittelt werden. Ein weiterer wichtiger, zentraler Bestandteil des Betriebssystemskerns ist die Kernel Library. Hier werden alle lebensnotwendigen Protokolle gespeichert, um die Kommunikation zu den wichtigsten peripheren Geräten sicherzustellen (bspw. DMA Controller, Timer, etc.). Diese Bibliotheken sind allerdings auch chipabhängig. Peripherie, die mit einer Usereingabe im Zusammenhang steht, kann wiederum in separaten DLLs gepackt werden - die sogenannten Kernel Extensions. Verschiedene Kernel Extensions können für Tastatur und viele mehr geschrieben werden. Die entsprechenden Extensions werden dem Symbian OS Image hinzugefügt, wobei der Kernel ihre Anwesenheit feststellt und beim Booten initialisiert (siehe Abbildung 3). Jeder Zugriff auf die Hardware wird über den Kernel geregelt. Die sonstigen Zugriffe werden ebenfalls durch den Kernel gesteuert.

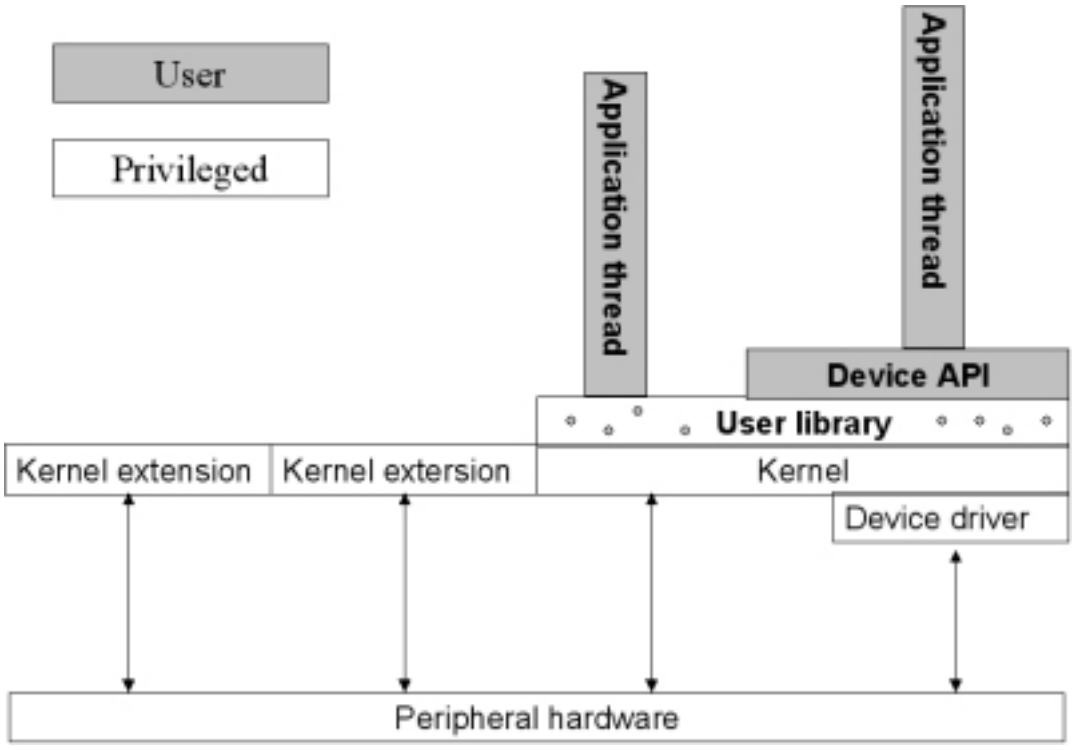

Abbildung 8.3: Der interne Aufbau des Betriebssystems.. [8]

Der Kernel selbst hat keinerlei Abhängigkeit von diesen Erweiterungen und auf die Kernel Extensions erhält keine Anwendung selbst den Zugriff. Damit nun die Anwendungen korrekt funktionieren und die für sie vorgesehene Hardware benutzen, sind just diese Informationen in den Treibern gespeichert. Diese können, um die Kontrolle über die Hardware für bestimmte Anwendungen zu erhalten, beliebig installiert und deinstalliert werden. Die Treiber bestehen aus zwei Teilen, einer Bibliothek, um die entsprechende API selbst anzusprechen und mehreren Bibliotheken, um die Hardware nutzen zu können. Damit kann die Kontrolle partiell vermittelt werden, ohne direkten Einfluss oder Schaden am Kernel beziehungsweise dem Betriebssystem selbst vorzunehmen. An dieser Stelle soll noch einmal näher auf die Kernel Side Library eingegangen werden. Es wurde ja bereits gesagt, dass hier alle lebensnotwendigen Daten gespeichert sind. Diese Bibliotheken unterteilen sich allerdings noch einmal in zwei verschiedene Typen. Zum einen gibt es den logical device driver DLL (LDD) und zum anderen den physical device driver DLL (PDD). Die logischen Treiber beinhalten sämtliche grundlegenden logischen Funktionen im Sinne von I/O. Diese sind auch immer auf andere Systeme transportierbar. Der interessante Aspekt sind die PDD, die jeweils auf das bestimmte Zielsystem zugeschnitten werden und die vorhandenen Ressourcen mobilisieren können. Diese Dateien sind notwendig und müssen verfügbar sein, aber durch die Zweiteilung muss man bei dem Portieren des Betriebssystems auf einer andere Hardware-Plattform nur die entsprechenden PDD austauschen und das System funktioniert in der gewünschten Umgebung. LDD und PDD sind hier als polymorphic DLL realisiert und müssen in einem bestimmten Verfahren genutzt werden. [8] Einzige Ausnahme aus der Prioritätscodierung ist der Screen Buffer. Das erscheint auch logisch, wenn man bedenkt, dass die Bildschirmdarstellung sich nicht “hinten anstellen“ sollte. Die Lösung liegt im Direct Memory Access, der die Daten unmittelbar in das LCD Display kopiert. Damit es an Schnelligkeit nicht mangelt, wurde der Screen Buffer mit allen notwendigen Lese- und Schreibberechtigung auf alle Prozesse ausgestattet. Damit werden Anwendungen über eine Graphic API schnell dargestellt. Folglich wird der privileged mode umgangen und dieser Vorgang liegt außerhalb der Prioritätencodierung.

8.4.2 Application-Sicht

Symbian OS wurde als ein umfangreiches Betriebssystem geplant, welches alle Fähigkeiten eines normalen Betriebssystem aufweist und dennoch in den Speicher eines Mobiltelefons passen sollte. [8] Das setzt bestimmte Anforderungen voraus. Diese wurden nun bereits genannt beziehungs-weise auch erläutert und teilweise auch diskutiert. Im Zusammenhang des Aufbaus wird nun die zweite Sichtweise betrachtet. Diese Sicht beschäftigt sich vor allem mit den Abhängigkeiten der Systemkomponenten. Es handelt sich allerdings um einen Übergang und eine klare Trennung zwischen den Sichtweisen kann nicht vorgenommen werden. So müssen zunächst noch einige tiefer gehende Erläuterungen vorangehen, bevor man sich dem allgemeinen Aufbau aus der Application-Sicht zuwenden kann.

Server-Client Struktur

Symbian OS beinhaltet eine umfassende Sammlung an Bibliotheken, um möglichst viele Industriestandards zu implementieren. Unter anderem TCP/IP, SSL, FTP und SMTP um nur einige aus verschiedenen Bereichen zu nennen. Abbildung 4 zeigt einige Bereiche auf,

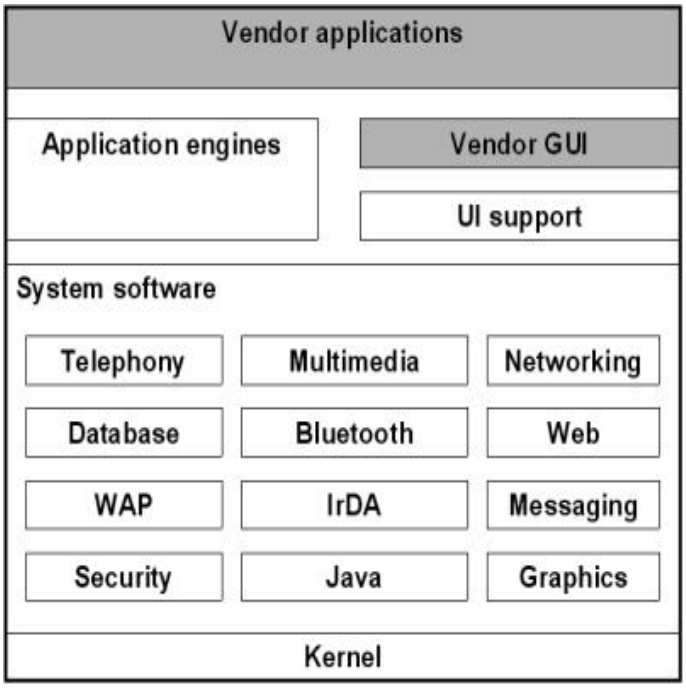

Abbildung 8.4: Dienste innerhalb der Symbian OS Technologies. [8]

die diese Standards innerhalb des Betriebssystems implementiert haben. Diese Dienste werden über eine einfache Server-Client Struktur eingebaut.

Der Server Thread läuft im unprivileged mode und jede Anwendung kann als Client über eine Standardschnittstelle mit einem Server durch den Kernel Kontakt aufnehmen. Dieser Bereich gehört zu dem bereits erwähnten Konzept der Prioritätencodierung. Die Nutzung dieser Client-Server Architektur in diesem Betriebssystem beinhaltet beispielsweise den file server und den media server. Anhand des media server folgt nun ein kleines Beispiel für eine Komponente (media codec), die durch das client-server System angesprochen wird. Abbildung 5 zeigt die Vermittlung zwischen zwei Anwendungen durch den Kernel. Der Kernel vermittelt zwischen beiden Komponenten und gewährt den Zugang zur Hardware.

Externe Anwendungen

Eine weitere Schicht in diesem Zusammenhang ist die der externen Anwendungen. Das beinhaltet unter anderem das Adressbuch, Editoren u.v.m.. Auf dieser Ebene ist auch das Grundgerüst für die GUI (Graphical User Interface) implementiert. Es ist bewusst keine komplette Oberfläche erstellt worden, um dem jeweiligen Hersteller eine individuelle Anpassung zu ermöglichen.

Weitere Dienste

Im Querschnitt werden nun weitere Dienstleistungen des Betriebssystem angeführt. Abbildung 6 verdeutlicht die Abstufungen untereinander, statt der detaillierten Schnittstellenlösungen. Die Abbildung soll dazu dienen, um eine komplette Übersicht über die Komponenten von Symbian OS 7.x zu erhalten. Man könnte vermuten, dass einige Komponenten nur integrierbar sind, wenn einige Subsysteme bereits vorhanden sind. Das stimmt nur bedingt und wird nach der genaueren Auflistung genauer erläutert. Auf der anderen Seite

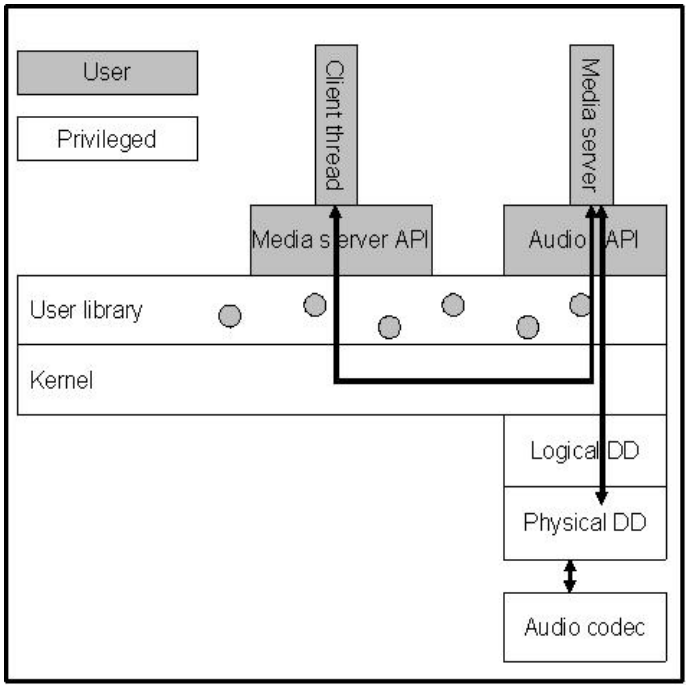

Abbildung 8.5: Der Zugriff auf eine Hardwarekomponente durch den media server. [8]

können auch hierarchisch, höhere Systeme installiert werden, wenn andere, die horizontal verschoben sind, nicht vorhanden sind.

Die Komponenten und ihre Subsysteme dieser Sichtweise sind folgende:

- **Base**

  Hier sind die grundlegenden Basiseinrichtungen implementiert. Dieser Bereich wurde bereits ausführlich in der Hardwaresicht erläutert. Bestandteile sind:

  Der Kernel und die Benutzerbibliotheken

  Die lebenswichtigen Peripheriegeräte

  Sicherheitssysteme

- **Telephony**

  Kommunikation wird im Sinne von Verbindungen über Weitverkehrsnetze verstanden. Mit Telefonie sind Anwendungen und der Zugriff auf anwendungsspezifische Systemparameter gemeint. Für Anwendungen aus der Telefonie stehen APIs zur Verfügung, die die grundsätzlichen Eigenschaften der unterschiedlichen zellularen Netze abstrahieren (GSM, GPRS, EDGE, CDMA als amerikanisches Äquivalent zu GSM, bzw. die Varianten von UMTS). Auf diese Weise lassen sich Anwendungen zwischen den unterschiedlichen Telefonstandards leichter portieren. Zu den gemeinsamen Eigenschaften gehören Informationen über Netze und über das Endgerät (Empfangsqualität, Ladezustand der Batterie, Kennungen der empfangenen Netze, Informationen über das jeweils genutzte Netz, Benachrichtigungen über Veränderungen, Kennung des Endgerätes), sowie der Zugriff auf Telefonverzeichnisse im Endgerät und in der SIM-Karte

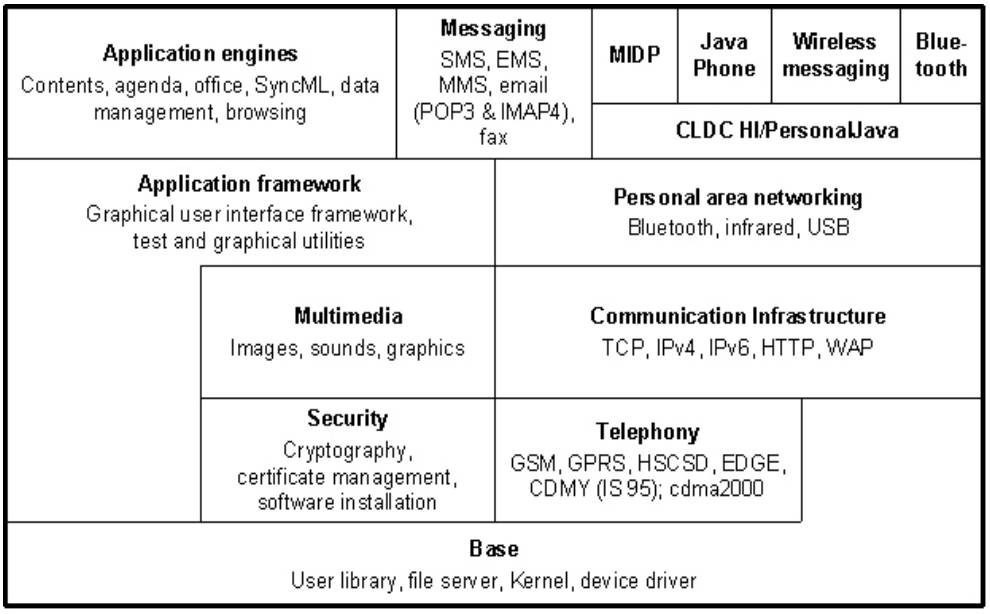

Abbildung 8.6: Struktur und Aufbau des Symbian OS 7.x [8]

- **Security**

  Stellt verschiedene Sicherheitsmechanismen zur Verfügung.

- **Application framework**

  In diesem Bereich sind die meisten wiederverwendbaren DLLs untergebracht, die sogenannte Middleware API. Middleware API sind Schnittstellen zur Bereitstellung von erweiterten Leistungen. Sie unterstützen Dateimanagement, Text, Grafiken und neben einigen anderen Features auch die Kern-GUI Komponenten.

- **Multimedia**

  Der media server stellt eine weitere, externe Komponenten dar, die dem Betriebssystem zugänglich gemacht wurde. Damit ist unter anderem die Aufnahme von Audiostücken möglich. Der media server wurde bereits im Zusammenhang mit der Server-Client Struktur erwähnt.

- **Communication infrastructure**

  Hier sind alle Protokolle wie TCP/IP, GSM, GPRS und WAP untergebracht. Weitere Kommunikationsmöglichkeiten sind Infrarot, Bluetooth und Seriell. Zur Kommunikationsinfrastruktur gehören ein Datenbankmanager für die Systemkonfiguration, Socket-Server und Schnittstellenmanager mit APIs als Gerüst für Kommunikations-nanwendungen, sowie Kommunikationsserver für serielle Verbindungen und für Verbindungen über WAP und HTTP. Die IP-Umgebung unterstützt sowohl IPv4 und IPv6 Adressen, sowie alle relevanten Protokolle aus dem IP-Stack (wie z.B. TCP, UDP, ICMP, PPP, TLS, SSL und Unterstützung für DNS). Ebenfalls vorhanden sind FTP, Telnet, sowie IPSec für VPN-Clients. Der HTTP-Stack wird z.B. für Anwendungen über GRPS verwendet, bzw. für SynchML oder die Übertragung von Streaming Medien über TCP/IP. Der WAP-Stack wird für Anwendungen über GSM, GPRS, CDMA und UMTS verwendet und unterstützt auch Push-Dienste.

- **Personal area networking**

  Zum Personal Area Network gehören der Bluetooth Protokollstack mit dem Generic Access Profil (GAP), dem Serial Port Profil (SPP) und dem Gerneral Object Exchange Profil (OPEX), sowie die Möglichkeit zur Vernetzung via Infrarot (IrDA) und USB.

- **Messaging**

  Zu den unter Messaging zusammengefassten Funktionen gehört die Unterstützung für SMS, EMS, MMS, E-Mail und FAX. Die Behandlung der unterschiedlichen Messaging-Formate ist grundsätzlich gleich. Eine EMS ist ein gegenüber der SMS erweitertes Format (Enhanced Messaging Service), das Bilder und Töne unterstützt und nach Art der SMS innerhalb der Mobilfunknetze übertragen wird. Eine MMS wird nach Art einer E-Mail über WAP oder HTTP transportiert. Die Messaging-Gruppe ermöglicht das Schicken von Nachrichten direkt aus anderen Anwendungen, wie z.B. das Versenden einer Visitenkarte (vCard) direkt aus dem Teilnehmerverzeichnis.

- **Application engines**

  Dieser Bereich deckt viele weitere Anwendungen ab, die als "nice-tohave" eingebaut wurden. Unter anderem zum Beispiel ein Terminkalender, To-Do Listen oder ähnliches.

- **CLDC**

  CLDC bedeutet so viel wie Connected Limited Device Configuration und beschreibt die Eigenschaften der Geräte im Low-End Bereich. Ausser diesen Konfigurationen, die generelle gerätespezifische Eigenschaften berücksichtigen, gibt es weitere Modelle für verschiedene Anwendungsklassen oder Anwendungsprofilen dieser Konfigurationen. Unter den Low-End Geräten unterscheidet sich das Profil eines Spielzeugautos von dem der Waschmaschine oder dem eines Mobiltelefons. Die APIs und Bibliotheken der Konfigurationen werden durch die APIs und Bibliotheken der Profile für den jeweiligen Anwendungszweck ergänzt. Für die Erstellung von Anwendungen auf Mobiltelefonen ist das Mobile Information Device Profile (MIDP) interessant. Die zugehörigen Anwendungen werden als Midlets bezeichnet.

- **Java MIDP**

  Vergleiche Kapitel 8.6 Java 2 Micro Edition (J2ME) - Java Umgebung für Embedded Anwendungen

Die verschiedenen Bereiche stellen Basiskomponenten dar, auf denen die Anwendungen wie beschrieben aufbauen beziehungsweise integriert werden können. Das Diagramm zeigt, bei den oben beschriebenen Komponenten eine gewisse horizontale, als auch vertikale Zugehörigkeit. [8] Es sei an dieser Stelle noch darauf hingewiesen, dass dieser getrennte Aufbau und die strikte Trennung der Anwendungen dazu führt, dass die Stabilität des Systems gewährleistet werden kann. Gleichsam kann das Betriebssystem voll operations- und multitaskingfähig sein.

8.5 Bewertung

Dieser Teil der Ausarbeitung beschäftigt sich mit dem Vergleich von Betriebssystemen und einer Bewertung der hier vorgestellten Systeme. Der Vergleich zu gängigen Betriebssystemen kann hauptsächlich durch eine Thematisierung erfolgen. Dabei würde man auf der Seite der gängigen Betriebssysteme das Schlagwort Generalisierung und auf der anderen Seite die Spezialisierung einführen. Gängige Betriebssysteme sind für viele, verschiedene Anwender und Plattformen gedacht. Es sollen möglichst zahlreiche Funktionen zugänglich sein, da anhand dieser der Marktpreis errechnet wird. Im Großen und Ganzen könnte man das zwar auch bei den Betriebssystemen für mobile Systeme sagen, allerdings ist es im Rahmen der Anforderungen nur bedingt möglich dem nachzukommen. Natürlich soll auch ein Mobiltelefon ein breites Anwendungsspektrum aufweisen, allerdings erwartet man von einem Handy beispielsweise keinen Raytracer. Übergreifend wurde bereits gezeigt, dass sich ein Betriebssystem für eine mobile Anwendung von einem gängigen Betriebssystem in den speziellen Anforderungen unterscheidet. Heutzutage werden aber in vielen Bereichen bereits normale Betriebssysteme in Abwandlungen oder als Basis genutzt, um sie in mobilen Endsystemen einzusetzen.

Vorausschauende Planung

Beginnend beim Aufbau unterscheidet sich dieser insofern, dass die hier vorgestellten Systeme besser durchdacht und geplant sind. Fehler sind in diesem Marktsegment unzulässig und enorme Anforderungen heben sie von den normalen Systemen ab. Vorausschauende Planung ist bei derartigen Betriebssystemen insofern notwendig, da die Entwicklung ständig fortschreitet und die einzelnen Systeme noch eine Kompatibilität aufweisen sollen.

Architektur

Insgesamt fällt die Architektur kleiner, aber auch geplanter aus. Das wiederum erkennt man auch beim Powermanagement. Ein fester Stromzugang ist meistens nicht möglich und ein großes Netzteil undenkbar. Die Auswahl von Bausteinen (Speicher, CPU, etc.) ist in mobilen Endprodukten begrenzter, allerdings von ihrer Bedeutung her der Hardware bei gängigen Systemen sehr ähnlich. Zumeist hat man nicht soviel Platz für zusätzliche Hardware und muss mit dem auskommen was vorhanden ist.

Kernel

Der Kernel in Betriebssystemen für mobile Systeme ist äußert kompakt und schnell ausführbar. In diesem Zusammenhang wurde auch die Ankopplung von zusätzlichen Anwendungen anhand Symbian OS gezeigt. Der Betriebssystemkern ist entsprechend 'Stand-Alone' ausgelegt und kann mit passenden Modulen ergänzt und aufgewertet werden.

Schnittstellen und Peripherie

Ein weiterer Bereich ist der der Schnittstellen. Während in gängigen Systemen Peripheriegeräte über beispielsweise das Plug & Play-System angesteuert beziehungsweise installiert werden, ist in dem Bereich mobiler Systeme zumeist nicht soviel Platz für Steckplätze für Erweiterungskarten. Die Ansteuerung von solchen Zusätzen, die wiederum klein ausfallen, muss sich hier größtenteils auf Softwareebene abwickeln lassen. Man will Platz und

unnötige Wege sparen. Notfalls kann das System allerdings auch gänzlich ohne weitere Anwendungen und Peripherie arbeiten. Falls in einem gängigen Betriebssystem eine wichtige Erweiterung fehlerbehaftet ist, kann es zu Abstürzen kommen. Gerade bei embedded Systemen ist, wie zuvor erläutert, ein Eingriff in das System nicht möglich und würde hier zu vorerst irreparablen Schäden führen.

Erreichbarkeit

Gängige Betriebssysteme können auf etablierte Protokolle zurückgreifen, während mobile Anwendungen grundsätzlich wireless bleiben müssen.

Echtzeit-Verhalten

Die Echtzeiteigenschaften sind bei vielen mobilen Anwendungen ein Muss, während gängige Betriebssysteme fast gänzlich darauf verzichten. Allerdings ist es mittlerweile laut einiger Hersteller 'problemlos' möglich ihr System mit Echtzeiteigenschaften auszustatten. [4]. Auch die neuesten Version von Symbian OS soll Echtzeiteigenschaften besitzen.

Im Zusammenhang auf die Bewertung interessieren folgende Aspekte:

- Kann man die Systeme miteinander vergleichen beziehungsweise welchen Sinn kann das haben?
- Welche Vor- und Nachteile beherbergen die einzelnen Systemen?

Um eine vernünftige Bewertung vorzunehmen, werden zunächst Bewertungskriterien benötigt. Bevor diese allerdings genauer diskutiert werden, sollte man sich vor Augen führen was man hier in eine Bewertung setzen will. Es soll der Versuch gestartet werden, die Systeme einander abzuwägen. Folgende Kriterien, die ähnlich des vorangegangenen Abschnitts von der Hardware-Sicht zur Application-Sicht gestaffelt sind, gelten als Bewertungsgrundlage:

- Kernel Größe
- Performance
- Portabilität
- Skalierbarkeit

8.5.1 Kernel

Als Bewertungsmaßstab in Diskussionen wird sehr gerne der Memory Footprint, also der Speicheraufwand eines Betriebssystems, herangezogen. An dieser Stelle hat Microsoft beispielsweise die Falschinformation verbreitet, dass sich bei Linux Embedded Systemen der Footprint im Bereich von mehreren Megabyte aufhalten würde. Dies trifft so aber nicht zu, da sich beispielsweise die Produkte der Firma LynuxWorks (BlueCat, LynxOS) in einem Bereich von ungefähr 250 KB aufhalten. Man muss sicherlich auch die Aussagen von LynuxWorks in diesem Zusammenhang kritisch betrachten, allerdings gibt es wiederum

andere Hersteller, die ihre Betriebssysteme ebenfalls auf Linux stützen und auf ihren offiziellen Internetpräsenzen ähnliche Ergebnisse verkünden. So hat zum Beispiel die Firma Trolltech das System Qt/EMBEDDED in ihrer Produktpalette und dieses begnügt sich mit rund 700 KB in der Minimalkonfiguration. Dahingehend liegt Symbian OS mit den bereits herangeführten 200 KB gut im Vergleich. Andere Firmen hingegen schweigen sich über ihre Minimalkonfigurationen aus.

8.5.2 Portabilität

Die Performance kann über Benchmark-Tests ermittelt werden. Das wäre auf der einen Seite eine gute Grundlage zur Bewertung der Systeme, aber auf der anderen Seite müsste man sich dann wiederum auf ein bestimmtes Zielsystem festlegen. Die Performance ist stark abhängig von dem Endprodukt und den damit verbundenen Anforderungen, so dass man vermutlich in jeder Kombination für das eine oder andere Betriebssystem argumentieren könnte. Fehlende Features sind aber wiederum eine gute Ausgangslage zur Bewertung. In diesem Zusammenhang sei auf die Echtzeiteigenschaften hingewiesen, die Symbian nicht aufweist und Microsoft auch nur bedingt leistet.

8.5.3 Skalierbarkeit

Die Skalierbarkeit bei einem Betriebssystem wie LynxOS beziehungsweise Windows Mobile 2003 ist aufgrund der Zielorientierung gegeben. Es soll an dieser Stelle nicht erneut auf den Vergleich zwischen den beiden Systemen hingewiesen, sondern der allgemeine Umstand betont werden, dass beispielsweise ein spezialisiertes System wie Symbian OS hier nicht in Konkurrenz stehen kann.

8.5.4 Performance

Die Performance kann über Benchmark-Tests ermittelt werden. Das wäre auf der einen Seite eine gute Grundlage zur Bewertung der Systeme, aber auf der anderen Seite müsste man sich dann wiederum auf ein bestimmtes Zielsystem festlegen. Die Performance ist stark abhängig von dem Endprodukt und den damit verbundenen Anforderungen, so dass man vermutlich in jeder Kombination für das eine oder andere Betriebssystem argumentieren könnte. Fehlende Features sind aber wiederum eine gute Ausgangslage zur Bewertung. In diesem Zusammenhang sei auf die Echtzeiteigenschaften hingewiesen, die Symbian nicht aufweist und Microsoft auch nur bedingt leistet.

8.6 Fazit

Es könnten an dieser Stelle noch viele weitere Kriterien als Bewertungsgrundlage gelten, allerdings fällt dann die Bewertung zunehmend schwerer. Viele Systeme beschränken sich auf

einen bestimmten Bereich und wollen nicht in Konkurrenz mit anderen Betriebssystemen stehen. Weitere Kriterien könnten aber beispielsweise Verwendete Bausteine, Unterstützte Features, Benutzerfreundlichkeit (in dem Zusammenhang eher Enwicklerfreundlichkeit), u.v.m. sein. Man kann erkennen, dass entweder die Grundlage oder die Ausrichtung selbst nicht immer gleich ist. Daher fällt auch ein abschließendes Urteil schwer. Es soll an dieser Stelle darauf verzichtet werden ein Präferenz zum Ausdruck zu bringen. Vor- und Nachteile wurden hier diskutiert und es sollte hervorgegangen sein, dass die beabsichtigte Anwendung Priorität bei der Auswahl eines solchen Betriebssystems hat. Der Ausblick am Ende dieser Ausarbeitung gibt in diesem Zusammenhang eine klarere Aufstellung und Abgrenzung für die Zukunftsaussichten in diesem Bereich.

8.7 Java 2 Micro Edition (J2ME) - Java Umgebung für Embedded Anwendungen

J2ME ist eine Teilmenge von J2SE, die das Minimum an Unterstützung beinhaltet, das für mobile Endgeräte benötigt wird. Die Aussage der Teilmenge muss gleich noch relativiert werden, aber für den Moment nehmen wir dies so an. Da es im Bereich der Kleingeräte eine so breite Palette an verschiedenen Hardwarevoraussetzungen gibt, entschied sich SUN für die Möglichkeit der gerätegruppenspezifischen Erweiterung der Spezifikation. Dies klingt erst einmal sehr kompliziert, bedeutet aber nichts anderes, als dass ein Teil der J2ME für alle Geräte festgelegt ist, wobei ein anderer Teil für eine entsprechende Gerätegruppe spezifisch ausgelegt ist. Diese spezifischen Auslegungen nennt man Konfigurationen und Profile. [2]

8.7.1 Einordnung und Zielsetzung von MIDP

Sun hat bei der Entwicklung von Java 2 eine Aufteilung in drei Teile (J2EE, J2SE, J2ME) vorgenommen. Welcher Teil jeweils benutzt wird, hängt im wesentlichen vom Einsatzgebiet, der zu erstellenden Applikation ab. Wir beschäftigen uns mit dem Einsatz von Java in Handys, Pagers und Smartphones. Die sogenannte Konfiguration, die in diesem Gebiet eingesetzt wird, trägt den Namen CLDC (Connected Limited Device Configuration). Da das API, das mit dieser Konfiguration ausgeliefert wird, dem Entwickler nur einige Grundfunktionalitäten zur Verfügung stellt, bedarf es einer Erweiterung dieses API's mittels eines zusätzlichen Profils. Das Profil wird in diesem Einsatzgebiete mit "Mobile Information Device Profile" (MIDP) bezeichnet.

8.7.2 Architektur

In diesem Abschnitt geht es um den architektonischen Aufbau der Software, auf welcher MID-Applikation resp. MIDlets ausgeführt werden können (siehe Abbildung 6). Auf der jeweiligen Hardware eines "Mobile Information Device" (MID) läuft ein relativ kleines

Betriebssystem, das die wichtigsten Hardware-Eigenschaften kapselt. Auf diesem Betriebssystem setzt die KVM (K Virtual Machine) auf. Die KVM übernimmt, wie wir es von einer Virtual Machine gewohnt sind, das Abbilden von Java-Bytecode-Instruktionen auf das Betriebssystem.

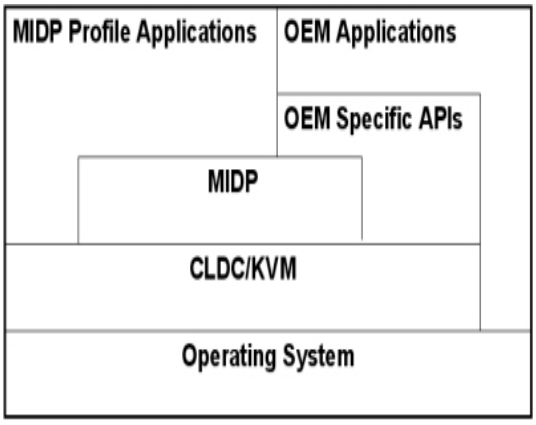

Abbildung 8.7: Position des Mobile Information Device Profiles (MIDP) innerhalb des Runtime Enivirnonment

Unmittelbar auf der KVM setzt das CLDC-API auf, welches seinerseits Dienste für das MIDP-API zur Verfügung stellt. Das MIDP-API bietet dem Entwickler eine Reihe von Funktionen an, die in der gleichnamigen Spezifikation definiert sind. Eine MID-Applikation, die ihrerseits auf dem CLDC-API resp. MIDP-API aufsetzt, sollte "theoretisch" plattform-unabhängig sein. Neben dem MIDP-API ist es dem Entwickler selber überlassen weitere Funktionen zu implementieren, welche möglicherweise auf ein spezifisches Gerät zugeschnitten sind (in Abbildung 2 als "OEM-Specific-API" bezeichnet). Die. Applikationen, die von solchen Funktionen Gebrauch machen oder möglicherweise direkt auf Betriebssystemfunktionen zugreifen, reizen vielleicht die spezifischen Dienste eines Gerätes besser aus, verlieren aber mit Sicherheit die Plattformunabhängigkeit.

8.7.3 Voraussetzungen

Die Variation an Betriebssystemen in MID-Geräten ist relativ groß. Einige Betriebssysteme bieten ein ausgereiftes Dateisystem an und andere stellen nichts derartiges zur Verfügung. Aufgrund der großen Unterschiede setzt die MIDP-Spezifikation die folgenden Voraussetzungen an das Betriebsystem eines MID voraus:

- **Kernel**
  Es wird nur ein kleiner Kernel verwendet, zur Hardwaresteuerung, Interruptbehandlung, für Exceptions und minimales Scheduling.
- **Storage**
  MIDP bietet für MIDlets die Möglichkeit Daten zu speichern und diese später wieder zu verwenden. Der persistente Speicher wird als Record Management System (RMS) bezeichnet und ist einfach als Datenbank realisiert. Die Daten werden als Records abgelegt.

- **Networking**
  Networking ist mit Hilfe einer Teilmenge des http-Protokolls realisiert. Es können sowohl IP-Protokolle als auch WAP oder i-mode genutzt werden.
- **Timers**
  CLDC bietet keine Timerfunktion. MID führt java.util.Timer und java.util.Timertask ein. Damit können MIDlets Timer nutzen. Applicationen können so Aktionen verzögern oder später ausführen.
- **Display**
  Bildschirmgröße: 96 x 54, Farbtiefe

8.8 Ausblick

Der Markt für Betriebssysteme für mobile Systeme ist äußerst umfangreich. Es gibt Unmengen an Systemen und Derivaten. Aber nach dem fünften Linux-Klon kommen nur noch wenig neue Aspekte hinzu. Im Rahmen einer Seminararbeit war es leider nicht möglich auf alle einzugehen und sämtliche Systeme vorzustellen. Auch wenn dieses Streben möglicherweise schon im Vornherein dank der Größe dieses Marktes zum Scheitern verurteilt sein könnte, so wäre es sicherlich im allgemeinen Interesse gewesen auf weitere Betriebssysteme, wie z.B. Palm OS einzugehen. Ein weiterer Aspekt ist der der Fächerung. Mobile Endsysteme gehen teilweise in ganz unterschiedliche Richtungen und sind nur auf unterster Ebene miteinander vergleichbar.
Es mag mobile Systeme geben, die nicht eingebettet sind (moderne PDAs, die Zugriff auf das OS zulassen); es gibt auch embedded Systems, die nicht mobil sind. Das ändert an den Anforderungen oder dem grundlegenden Aufbau dieser Systeme kaum etwas. Die Anforderungen sind klar und der Aufbau wurde erläutert. Wichtig ist das Einsatzgebiet. Wie schon erwähnt, ist der Markt für mobile Endsysteme sehr umfangreich und groß. Allein der Anteil der embedded systems ist enorm. Und auch die Entwicklungen schreiten stetig voran, so dass ein Ausblick nur unter den aktuellen Begebenheiten gültig sein kann und damit wahrscheinlich schon in naher Zukunft überholt sein könnte. Man setzt zunehmend auf übergreifende Dienstleistungen, die dem Benutzer das Leben vereinfachen sollen. Und diese Entwicklung passiert nicht nur in diesem Marktsegment, sondern durchdringt alle Hersteller von mobilen Endprodukten. Multi-Funktionalität wird durch einen zunehmenden Satz an Anwendungssoftware gefördert (vgl. SmartPhones, Handy mit Kamera, etc.). Abseits dieser Entwicklung herrscht immer noch ein harter (ggf. auch ideologischer) Konkurrenzkampf zwischen Windows und Linux Anhängern statt. Wer sich hier behaupten kann, wäre zu diesem Zeitpunkt kaum abzuwägen. Es gibt aktuell zu viele Argumente für und wider einem bestimmten Betriebssystem, so dass man höchstens die Vermutung aufkeimen lassen könnte, dass die Firmen mit dem "längerem Atem" das Rennen machen. So gibt es zwar eine hohe Bandbreite an Systemen, aber im Falle Microsoft hat die Firma eindeutig höhere Mittel, um ihre Produkte auf dem Markt zu forcieren. Symbian OS hat derweil ihre 7.x Reihe herausgebracht. Ebenso Einfluss, Marktgröße und Erfahrung können nur bedingt als Erfolgsgarant für die Zukunft gelten. Fakt ist allerdings, dass dieser Markt einen hohen Einfluss auf das alltägliche Leben hat und immer mehr Zuwachs erhält.

Literaturverzeichnis

[1] GNU Project, 2003, http://www.gnu.org/ http://www.gnu.org/copyleft/gpl.html

[2] Java 2 Plattform Micro Edition, http://java.sun.com/j2me/ http://java.sun.com/products/midp/

[3] LynxOS-178 RTOS for Software Certification, Lynux- Works, White Papers, 2003, http://www.lynuxworks.com/products/ Linux on Mobile Computers, http://www.tuxmobile.org

[4] Microsoft Windows, http://www.microsoft.com/windowsmobile/

[5] Schiffmann W., Schmitz R.: Technische Informatik - Grundlagen der digitalen Elektronik, Springer Verlag, Heidelberg, 4. Auflage, 2001

[6] SCO OpenServer - Documentation Library (SCO Open- Server Handbook), Technical Publication, 2003, http://osr5doc.ca.caldera.com

[7] Sony - Glossary 2003 http://www.css.ap.sony.com/Glossary /

[8] Symbian OS, The Mobile Operating System, http://www.symbian.com

[9] Witzak, M.P.: Echtzeit Betriebssysteme - Eine Einführung in Architektur und Programmierung, Franzis Verlag, Poing, 2000